11-30-00 A

AVIRON
297 N. Bernardo Avenue
Mountain View, California 94043
650.919.6637

Atty. Docket No. AVIR-0032-1

"Express Mail" Label No. EF337287150US

Date of Deposit 28 November 20000

PATENT APPLICATION
ASST. COMMISSIONER FOR PATENTS
Washington, D. C. 20231

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Asst. Commissioner for Patents, Washington, D. C. 20231

By Tracy J. Dunn

Sir:

Transmitted herewith for filing is the
[x] patent application of
[ ] design patent application of
[ ] continuation-in-part patent application of

Inventor(s): Kemble G.W., Duke, G.M., and Spaete, R.R.

For: ATTENUATION OF CYTOMEGALOVIRUS VIRULENCE

[ ] This application claims priority from each of the following Application Nos./filing dates: _____/_____; _____/_____; _____/_____.

[ ] Please amend this application by adding the following before the first sentence: --This application claims the benefit of U.S. Provisional Application No. 60/_____, filed _____.--

Enclosed are:

[x] 36 sheet(s) of [ ] formal [ ] informal drawing(s).
[ ] An assignment of the invention to _____ .
[x] A [ ] signed [x] unsigned Declaration & Power of Attorney.
[ ] A [ ] signed [ ] unsigned Declaration.
[ ] A Power of Attorney.
[ ] A verified statement to establish small entity status under 37 CFR 1.9 and 37 CFR 1.27 [ ] is enclosed [ ] was filed in the earliest of the above-identified patent application(s).
[ ] A certified copy of a _____ application.
[ ] Information Disclosure Statement under 37 CFR 1.97.
[ ]

In view of the Unsigned Declaration as filed with this application and pursuant to 37 CFR §1.53(d), Applicant requests deferral of the filing fee until submission of the Missing Parts of Application.

DO NOT CHARGE THE FILING FEE AT THIS TIME.

Tracy J. Dunn
Reg. No.: 34,587
Attorney for Applicant

Telephone:
650.919.6637

APPNOFEE.TRN 4/96

Aviron Docket No. AVIR-0032-1

PATENT APPLICATION

ATTENUATION OF CYTOMEGALOVIRUS VIRULENCE

Inventor: George W. Kemble, a United States citizen residing in Fremont, California

Gregory M. Duke, a United States citizen residing in Palo Alto, California

Richard R. Spaete, a United States citizen residing in Redwood City, California

Assignee: AVIRON

PATENT
AVIR-0032

ATTENUATION OF CYTOMEGALOVIRUS VIRULENCE

TECHNICAL FIELD

The present invention is related generally to methods and compositions for treating or preventing cytomegalovirus (CMV) infections, such as congenital CMV disease, CMV retinitis, CMV mononucleosis, and the like, and methods of attenuating pathogenic cytomegalovirus isolates and strains, genetically engineered cytomegaloviruses and combinations thereof, methods for altering the phenotype of CMV viruses, attentuated viral vaccine compositions, and uses thereof. More particularly, the present invention is related to methods and compositions for prophylaxis and therapy of human cytomegalovirus infection, including the use of methods that functionally inactivate a subset of cytomegalovirus genes present in pathogenic isolates of human cytomegalovirus.

BACKGROUND

Cytomegalovirus (CMV) is a widespread herpesvirus in the human population, with between 0.2 and 2.2% of the infant population becoming infected in utero and another 8-60% becoming infected during the first six months of life (Reynolds et al. (1973) New Engl. J. Med. 289: 1). Although CMV infections are most commonly subclinical, CMV-induced sensorineural hearing loss and fatal cytomegalovirus infections ("cytomegalic inclusion disease") are important public health problems. Moreover, CMV is one of the more common opportunistic infections associated with Acquired Immune Deficiency Syndrome ("AIDS") and frequently produces disease, with recurrent infection occurring in HIV-positive individuals, typically taking the form of retinitis or ulcerative lesions in the colon and esophagus, and occasionally producing extensive necrotization of the bowel with a grave prognosis (Rene et al. (1988) Dig. Dis. Sci. 33: 741; Meiselman et al. (1985) Gastroenterology: 88: 171). Cytomegalovirus (CMV) infection is the major infectious cause of mental retardation and congenital deafness. CMV is also responsible for a great deal

of disease among the immunosuppressed, producing general and often severe systemic effects in patients with AIDS, in organ transplant recipients who have been iatrogenically immunosuppressed, and in bone marrow transplant patients.

It is clear that cytomegalovirus infections are a significant human health problem. Therefore, it is desirable to develop prophylactic and therapeutic methods and compositions to prevent cytomegalovirus infection and/or inhibit recurrent infectious outbreaks from persistent latent infections, particularly for treating CMV retinitis, CMV mononucleosis, and related CMV pathology in human patients.

One approach that has been used to treat herpesvirus infections is to inhibit CMV viral DNA replication. For example, viral DNA replication can frequently be inhibited by agents that inhibit virally-encoded DNA polymerase. The most notable examples of such inhibitors of viral DNA polymerase are acyclovir, ganciclovir, citrusine-I, and the acyclic guanosine phosphonate (R,S)-HPMPC (Terry et al. (1988) Antiviral Res. 10: 235; Yamamoto et al. (1989) Antiviral Res. 12: 21). However, these compounds are not completely selective for viral thymidylate synthetases or DNA polymerases and therefore can disadvantageously cause inhibition of host DNA replication at high doses. Moreover, the development of mutant viruses which are resistant to the inhibitory effects of these compounds have been reported, and appear to result from mutations in the viral DNA polymerase (Coen et al. (1982) J. Virol. 41: 909; Coen et al. (1980) Proc. Natl. Acad. Sci. (U.S.A.) 77: 2265; Larder et al. (1987) EMBO J. 6: 169). Thus, while CMV infections, such as CMV retinitis, can be initially treated with foscarnet and ganciclovir, after a period of time CMV replication and progression of the pathological viral infection recurs.

Passive immunization with antibodies (e.g., immune globulin) has been tested in combination with ganciclovir for therapeutic efficacy in humans. Such antibody preparations are obtained from the serum of donors, who possess a high antibody titre to the virus as a result of an earlier infection. One disadvantage of such conventional antibody preparations is the

limited number of suitable donors and the poor reproducibility or quality of the various preparations, including potential contamination with pathogens and pathogenic viruses. Unfortunately, the use of intravenous immune globulin in combination with ganciclovir apparently does not produce significantly improved efficacy as compared to ganciclovir treatment alone (Jacobson et al. (1990) Antimicrob. Agents and Chemother. 34: 176).

The safety and pharmacokinetic profiles of anti-cytomegalovirus monoclonal antibodies are discussed in Aulitzky et al. (1991) J. Infect. Dis. 163: 1344 and Drobyski et al. (1991) Transplantation 51: 1190. However, none of the reported human anti-CMV monoclonal antibodies have been shown to possess significant therapeutic efficacy in treating CMV infections (e.g., retinitis) in humans.

Attempts to use recombinantly produced hCMV glycoproteins as a subunit vaccine to provide protective immunity against hCMV infection and pathogenesis have not proven to be effective, but remain candidates for additional evaluation.

Thus, there exists a need in the art for effective methods and compositions for inhibiting human cytomegalovirus replication, attenuating CMV virulence in vivo, neutralizing CMV virions, and for preventing and treating human cytomegalovirus infections, and especially CMV infections in preborns, newborns, and immunosuppressed patients such as AIDS patients. For example but not limitation, a suitable attenuated human CMV vaccine which elicits satisfactory immunoprotection against CMV infection is needed in the art. The present invention fulfills these and other needs.

SUMMARY OF THE INVENTION

A basis of the present invention is the surprising and unexpected finding that: (1) clinical isolates of pathogenic CMV variants contain a genomic region ("virulence region") which typically is not present in CMV strains which have undergone extensive laboratory passaging of the virus in cell culture (hereafter termed "highly passaged strain variants") and (2)

functional disruption (e.g., deletion or insertional inactivation and the like) of genes in this genomic region produces a substantial attenuation of CMV virulence and/or pathogenicity in vivo. Furthermore, the virulence region of a clinical isolate of CMV is frequently deleted, rearranged, or substantially changed over the course of passaging the virus in cell culture.

In one aspect of the invention, the virulence region is obtained from an early passage Toledo strain and is conveniently termed the "Toledo genomic region" herein, although equivalent (e.g., homologous) regions or subsequences thereof are present in other clinical isolates of CMV besides the Toledo strain of CMV; the term "Toledo genomic region" encompasses these homologous regions in other clinical CMV isolates, many early passage CMV strains, and non-isolated pathogenic CMV variants.

The Toledo genomic region which is present in pathogenic CMV isolates and which is typically substantially absent in highly passaged CMV strains (e.g., AD169, high-passage Towne) has been sequenced and several open-reading frames have been identified (PCT Publication WO96/30387, U.S.S.N. 08/414,926, U.S.S.N. 08/644,543 filed 10 May 1996, each incorporated herein in their entirety by reference). Functional disruption of these open reading frames, either singly or in combination, has been unexpectedly found to substantially reduce virulence of the resultant CMV mutant(s) in vivo. Thus, in part, the invention provides methods and compositions for suppressing or inactivating expression of genes of the Toledo genomic region and its homolog regions in other CMV variants, and thereby reducing virulence and pathogenicity of clinically important CMV variants to generate a "Toledo region-attenuated CMV variant"; such Toledo region-attenuated CMV variants have altered phenotypes which generally make them candidates for use in live attenuated virus vaccines for prophylaxis and/or treatment of CMV disease. The invention is, in part, further based on the heretofore unrecognized finding that pathogenic clinical isolates of CMV have a distinct genome as compared to the commonly used laboratory-passaged strains of human CMV (e.g., AD169, highly-passaged Towne), and that the genomic region which is present in the clinical isolates and

which is substantially absent in laboratory-passaged strains confers enhanced virulence in vivo. Most common approaches to development of CMV therapies and vaccines have heretofore relied on laboratory-passaged strains which typically lack all or part of the Toledo genomic region and the genes encoded therein which have been unexpectedly found to confer enhanced in vivo virulence and are believed to contribute to clinical pathology and CMV-related disease.

The invention provides a method for attenuating virulence of CMV comprising functionally inactivating at least one open reading frame in a virulence region of a CMV genome having substantial identity to at least 300 bp, typically at least 500 bp, of a 15 kb sequence present in the genome of the Toledo strain of CMV and absent from the genome of the AD169 strain of CMV and/or absent from the genome of highly-passaged Towne (i.e., more than 50-100 passages). In an aspect, the method functionally inactivates at least one open reading frame present in a genomic region of a CMV genome having substantial identity to at least 300 bp of a 13 kb sequence present in the genome of the Toledo strain of CMV and absent from the genome of the Towne strain of CMV. In an embodiment, the method functionally inactivates at least one open reading frame present in a genomic region of a CMV genome having substantial identity to at least 500 bp of the sequence shown in Figs. 1A through 1T. In an embodiment, the method functionally inactivates at least the open reading frame corresponding to UL148 as identified herein. In a variation, the method funtionally inactivates open reading frames in the region spanning UL138 to UL148. In an embodiment, the method functionally inactivates UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148. In a variation, UL148 is inactivated singly or in combination with other open reading frames of the Toledo genomic region. In a specific embodiment, UL148 is inactivated in combination with UL141 and/or UL144. Typically, such Toledo region-attenuated CMV variants comprise at least 500 bp of the Toledo genomic region or a homolog region having at least 80 percent sequence identity; frequently they comprise at least 1.0

kbp of the Toledo genomic region or homolog virulence region; often they contain at least 5.0 kbp to 8.0 kbp of the Toledo genomic region or homolog virulence region, and can comprise up to a complete Toledo genomic region or homolog virulence region. It is possible for a synthetic virulence region to be comprised of portions of two or more virulence regions (e.g., such as a chimeric virulence region comprising part of the Toledo genomic region from a first clinical isolate with a complementing portion of the Toledo genomic region of a second clinical isolate).

In an aspect, the invention provides a method for attenuating a CMV strain or isolate containing an encoding polynucleotide sequence encoding a polypeptide which is at least 80 percent sequence identical to a polypeptide encoded by UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148 of the Toledo genomic region; the method comprising functionally inactivating (e.g., deleting or introducing a nonsense or missense mutation) said encoding polynucleotide sequence to produce a Toledo region-attenuated CMV variant. In a variation, all open reading frames (ORFs) in the CMV isolate that are at least 80% sequence identical to the corresponding sequence of the Toledo genomic region are functionally inactivated. In a variation, all open reading frames (ORFs) in the CMV isolate that are at least 80% sequence identical to UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148 of the Toledo genomic region are functionally inactivated. In an alternate variation, only one or a subset of the open reading frames (ORFs) in the CMV isolate that are at least 80% sequence identical to the corresponding sequence(s) of the Toledo genomic region are functionally inactivated. Such Toledo region-attenuated CMV variants comprise at least 500 bp of a Toledo genomic region and can comprise up to a complete Toledo genomic region (including a chimeric Toledo genomic region composed from distinct clinical isolates or strains).

In an aspect, the invention provides a recombinant CMV virus, comprising a genome having at least 500 bp of a virulence region wherein at least one ORF has been functionally inactivated by a genetic alteration which is predetermined and/or which does

not occur in known isolates or strains of CMV regardless of passage history.

In an aspect, the method of attenuating virulence comprises functional inactivation of open reading frames by predetermined structural mutation (e.g., deletion, insertion, missense or nonsense mutation, and the like) of at least one open reading frame, or a predetermined mutation of a transcriptional control sequence that controls transcription of the open reading frame, or predetermined mutation of a splicing signal sequence or the like necessary for efficient expression of the encoded gene product of the open reading frame. In an embodiment, a selectable marker gene is introduced into an open reading frame, often in the portion of the open reading frame believed to encode the amino-terminal two-thirds of the gene product, to structurally disrupt the open reading frame and result in the inactivation of the open reading frame's capacity to encode its functional gene product. In a variation, open reading frame UL148 is structurally disrupted by mutation; in one embodiment the structural disruption results from insertion of a selectable and/or screenable marker gene (e.g., *gpt/lacZ*). In an embodiment, a selectable marker gene is used to replace all or part of at least one open reading frame, such as by replacement of a deleted region of the Toledo genomic region with a selectable marker gene. In a variation, a region spanning open reading frame UL138 to UL148 is structurally disrupted by mutation; in one embodiment the structural disruption results from deletion of the UL138-UL148 region and replacement with a selectable and/or screenable marker gene (e.g., *gpt/lacZ*).

In an aspect, the functional inactivation of a Toledo genomic region gene is provided by transcriptional and/or translational suppression with an antisense polynucleotide having a sequence of at least 15 nucleotides, typically at least 25 nucleotides, that are substantially complementary to a Toledo genomic region, most usually the antisense polynucleotide is substantially complementary to an open reading frame sequence of a Toledo genomic region open reading frame. In an embodiment, the antisense polynucleotide is substantially complementary to

at least 25 nucleotides of UL148. In an embodiment, the antisense polynucleotide is complementary to UL148 and further comprises additional 5' and/or 3' nucleotide(s) which are not substantially complementary to UL148. In variations, the antisense polynucleotides comprise non-natural chemical modifications, and can include, for instance, methylphosphonates, phosphorothioates, phosphoramidites, phosphorodithioates, phosphorotriesters, and boranophosphates. In a variation the antisense molecules can comprise non-phosphodiester polynucleotide analogs wherein the phosphodiester backbone is replaced by a structural mimic linkage include: alkanes, ethers, thioethers, amines, ketones, formacetals, thioformacetals, amides, carbamates, ureas, hydroxylamines, sulfamates, sulfamides, sulfones, and glycinylamides. In a variation, the invention provides peptide nucleic acids (PNAs) having a nucleobase sequence which is substantially complementary to a Toledo genomic region sequence, such as an open reading frame (e.g., UL148, UL141, UL142, etc.).

The invention also provides attenuated live virus CMV vaccines wherein at least one open reading frame of a Toledo genomic region is structurally disrupted. Typically, the UL148 open reading frame is structurally disrupted, either singly or in combination with other Toledo region open reading frames (e.g., UL141, UL144, and the like). Often the disruption of the open reading frame is an insertion, deletion, or replacement mutation which confers the property of reduced virulence as determined by a suitable in vivo virulence assay (e.g., see Experimental Examples). Toledo genomic region mutants which exhibit at least one log reduction, preferably two logs or more reduction, in virulence as determined by in vivo virulence assay, or other equivalent virulence measure, are attenuated CMV vaccines. Such attenuated CMV vaccines are used to immunize individuals to confer protective immunity, typically antibody-mediated and/or cell-mediated immunity, to prevent or reduce the severity of subsequent CMV infection following a suitable immunization period.

In an aspect, the invention also provides attenuated

live virus CMV vaccines wherein at least one open reading frame of a Toledo genomic region is replaced by a segment of the Towne genome which is not present in AD169. The Towne genome comprises a region not present in AD169; the region contains open reading frames designated UL147, UL152, UL 153, and UL154 and generally is spanned by nucleotides 178221 to 180029 of the Towne genome according to the AD169 numbering convention. An attenuated virus of the invention can, in one embodiment, comprise a Toledo genome wherein the Toledo genome region spanning open reading frames UL133 to UL151 are replaced with a Towne genome region spanning UL147, UL152, UL153, and UL154; this engineered CMV virus variant is an attenuated Toledo virus which comprises desirable features of Towne while reducing undesirable virulence of the Toledo genome region. The invention provides other variations of this basic method, whereby a segment of the Toledo genome region comprising at least one open reading frame is deleted or otherwise structurally disrupted in a CMV variant having a Toledo genome region or its homolog, and a segment of a Towne genome region comprising at least one open reading frame in inserted in the CMV variant. In an embodiment, the engineered CMV variant comprises: (1) Toledo DNA (DNA substantially identical to a Toledo strain, preferably identical to it) from about nucleotides 1 to about 168,000 corresponding to (i.e., according to) the AD169 nucleotide numbering convention, operably linked to (2) Towne DNA (DNA substantially identical to a Towne strain, preferably identical to it) from about nucleotides 143,824 to 189,466 according to the AD169 nucleotide numbering convention, operably linked to (3) Toledo DNA (DNA substantially identical to a Toledo strain, preferably identical to it) from about nucleotides 189,466 to about 209,514 corresponding to (i.e., according to) the AD169 nucleotide numbering convention, operably linked to (4) Towne DNA (DNA substantially identical to a Towne strain, preferably identical to it) from about nucleotides 200,080 to 229,354 according to the AD169 nucleotide numbering convention. The invention also provides vaccine compositions and formulations of such attenuated CMV viruses, which can include adjuvants, delivery vehicles, liposomal formulations, and the

like. The invention also provides the use of such attentuated CMV variants for prevention of CMV disease and infection; in one aspect this use includes administration of such vaccine to human subjects.

In a variation, the functional inactivation of a Toledo genomic region gene is provided by suppressing function of a gene product encoded by a Toledo region open reading frame by contacting or administering an antibody which is specifically reactive with said gene product. In an embodiment, the Toledo genomic region gene is UL148, UL141, and/or UL144, typically at least UL148, although other Toledo open reading frames can be used. The antibody binds to a gene product encoded by a Toledo region open reading frame with an affinity of at least about $1x10^7$ M^{-1}, typically at least about $1x10^8$ M^{-1}, frequently at least $1x10^9$ M^{-1} to $1x10^{10}$ M^{-1} or more. In some aspects, the antibody is substantially monospecific. In an embodiment, the antibody is a human antibody raised by immunizing an individual with an immunogenic dose of a gene product of a Toledo region open reading frame. In an embodiment, the human antibody is a monoclonal antibody, or collection of human monoclonal antibodies which bind to the Toledo region gene product(s). In an embodiment, the antibody is a humanized antibody comprising complementarity-determining regions substantially obtained from a non-human species immunoglobulin reactive with the Toledo region gene product, and further comprising substantially human sequence framework and constant regions. The invention also comprises pharmaceutical formulations of such antibodies and the use of such antibodies to treat or prevent CMV diseases, such as by passive immunization or the like.

In an aspect, the invention provides a composite CMV variant comprising a highly-passaged Towne genome and at least one open reading frame of a Toledo genome region, typically present in or adjacent to the U_L/b' region of the composite CMV. In an aspect, the composite CMV is a highly-passaged Towne genome further comprising a Toledo UL148, UL141, and/or UL144. In an embodiment, the composite CMV is a highly-passaged Towne genome with a complete Toledo genome region; in a variation said Toledo

genome region has at least one open reading frame functionally inactivated to further attenuate the virulence of the composite CMV. In a variation, a low passage Towne genome (i.e, less than 40 passages in culture) is used in place of a highly-passaged Towne genome. In an alternate variation, a virulence region from a low-passage Towne genome is emplaced in a Toledo genome so as to thereby replace at least 1 kpb of the virulence region of the Toledo genome with at least 500 bp, typically approximately the same length, of a corresponding region (e.g., substantial sequence identity) of low-passage Towne.

In an aspect, the invention provides a chimeric CMV virus, comprising a genome having a plurality of polynucleotide sequences, linked in conventional phosphodiester linkage, wherein at least two of said polynucleotide sequences are derived from different clinical isolates or strains of CMV. Said chimeric CMV virus can comprise a genome having a plurality of polynucleotide sequences, linked in conventional phosphodiester linkage, wherein a first CMV genome sequence of at least 500 bp and less than a complete CMV genome length (e.g., less than 250 kbp) is at least 98 percent sequence identical to a first CMV isolate or strain, and at least one additional CMV sequence of at least 500 bp and less than a complete CMV genome length (e.g., less than 250 kbp) is at least 98 percent sequence identical to a second CMV isolate or strain which has a genome having a polynucleotide sequence of at least 500 bp which is less than 60 percent sequence identical to any portion of the genome of said first CMV isolate or strain and/or which is absent or substantially absent in the genome of said first CMV isolate or strain. Said chimeric CMV virus comprises a genome having sufficient genetic information to replicate as a virus, typically as an infectious virus, in suitable host cells or a suitable host organism or replication system (e.g., SCID/hu thy/liv mice, human lung fibroblasts, and other systems known inthe art). Generally, said chimeric CMV virus has a genome that comprises genetic information which is substantially sequence identical, generally at least 80 percent sequence identical, usually at least 95 percent sequence identical or more, to a high-passage Towne genome; said chimeric

CMV virus genome typically further comprises genetic information which is substantially sequence identical, generally at least 80 percent sequence identical, usually at least 95 percent sequence identical or more, to at least 1 kbp of a virulence region of a clinical isolate of CMV or a low-passage strain of CMV other than low-passage Towne; in an embodiment, a complete virulence region (e.g., Toledo genome region) of a clinical isolate or low-passage CMV strain is present.

In an aspect, the invention provides a chimeric CMV virus, comprising a chimeric genome comprising a polynucleotide having a first CMV sequence of at least 500 bp having at least 97 percent sequence identity with a genome of a first CMV isolate or CMV strain and a second CMV sequence of at least 500 bp having at least 97 percent sequence identity with a genome of a second CMV isolate or CMV strain, and wherein said chimeric genome comprises genetic information having substantial identity (e.g., at least 80 percent sequence identity, preferably at least 95 percent sequence identity) spanning at least about the complete low-passage Towne genome. Typically, the chimeric genome comprises at least 500 bp containing at least one ORF having at least 95 to preferably 100 percent sequence identity to a virulence region (e.g., Toledo genome region) of a clinical isolate or low-passage strain of CMV other than low-passage Towne.

In an aspect, the invention provides a chimeric CMV virus, comprising a chimeric genome comprising a polynucleotide having a first CMV sequence of at least 500 bp having at least 97 percent sequence identity with a genome of a first CMV isolate or CMV strain and a second CMV sequence of at least 500 bp having at least 97 percent sequence identity with a genome of a second CMV isolate or CMV strain, and wherein said chimeric genome comprises genetic information having substantial identity (e.g., at least 80 percent sequence identity, preferably at least 95 percent sequence identity) spanning at least about the complete Toledo genome excepting at least 1 kbp of the virulence determinign region of Toledo (Toledo genome region), and preferably excepting at least 5 kbp to the entire approximately

15kbp virulence-detemrining Toledo genome region. Typically, the chimeric genome comprises at least 500 bp containing at least one ORF having at least 95 to preferably 100 percent sequence identity to a virulence region of low-passage Towne.

In specific embodiments, the invention provides exemplary CMV chimeric viruses composed of genome portions of high-passage Towne and genome portions of Toledo; the exemplary CMV chimeric viruses are designated herein as Chimera I, Chimera II, Chimera III, Chimera IV, and Towne/Tol 11. In an aspect, the invention encompasses these specific embodiments and variants of each exemplified Chimera wherein the boundaries (splice junctions/recombination joints) between the various Towne and Toledo genome portions vary from the specific exemplified Chimeras by less than 20 kbp, typically less than 10 kbp, usually by less than 5 kbp, and in many embodiments by less than 1 kbp from the specific examples provided herein.

In a variation, the invention provides a diagnostic method for identifying a virulent CMV strain in a sample by detecting the presence of unique Toledo genome region polynucleotide sequences and/or by detecting the presence of a polypeptide encoded by an open reading frame of the Toledo genomic region. Detection of polynucleotide sequences can be by any suitable method, including but not limited to PCR amplification using suitable primers, LCR, hybridization of a labeled polynucelotide probe, and the like. Detection of polypeptide speceis is typically done by immunoassay using a pecific antibody to the Toledo region gene product(s).

The invention also provides a method of treating or preventing CMV infection, the method comprising administering to an individual an efficacious dose of a polypeptide which is substantially identical to the deduced amino acid sequence of UL148. In a variation, the polypeptide is a truncated variant, mutein, or analog of the deduced amino acid sequence of UL148, wherein the polypeptide is soluble.

A further understanding of the nature and advantages of the invention will become apparent by reference to the remaining portions of the specification and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1R. Nucleotide sequence of Toledo genome region isolated from Toledo strain of HCMV.

Figures 2A-2H. Deduced amino acid sequences of open reading frames UL130 through UL151. Convnetional single letter abbreviations are used.

Figure 3. Schematic representation of open reading frames and their location in Toledo genome region. Top line schematically portrays entire Toledo genome with U_L/b' region identified. Bottom line shows enlarged view of U_L/b' region. Arrows indicate polarity and length of open reading frame. Solid circles indicate potential glycosylation sites.

Figure 4. Schematic comparison of the novel genome regions of Toledo and highly-passaged Towne as compared to AD169.

Figure 5. CMV Towne and Toledo cosmids used to regenerate specific chimeric CMV viruses. The location of the cosmid insert are indicated beneath the appropriate viral genome. The numbers at the end of the insert denote the endpoints determined by DNA sequence analysis; the numbers correspond to AD169 genomic sequence in GenBank. "XXX"" denotes an end which was refractory to DNA sequence analysis. These ends were mapped by restriction enzyme and Southern blot analyses. The vertical dashed line represents the location of the internal "a" sequence of the virus. The lower line depicts the structure of the Tol/Twn 39/50 genome. The thick gray line denotes sequences derived from Toledo and the thin black line depicts sequences contributed from highly-passaged Towne strain. Regions of overlap could be derived from either virus and are represented by a region of a thick gray and a thin black line together. The Tol/Twn 39/50 genome does not contain the Toledo genomic region.

Figure 6. Analysis of the gpt/LacZ recombinant viruses in the SCID-hu (thy/liv) model. Two independent isolates of Tol pGD6 and Tol pGD7 were tested in the model. 3 mice were used per group and the mean of the data is displayed. Error bars representing 2 standard errors from the mean are also displayed.

Figure 7. Southern blot showing that a variety of clinical isolates of CMV contain sequences homologous to the

Toledo U_L/b' region. The Towne lane contains genomic DNA from Aviron's highyl-passaged Towne strain (Towne AV).

Figure 8. Southern blot showing that previous variants of the Towne strain hybridize to the Toledo U_L/b' region. Twn•Merck indicates Towne strain from the Merck clinical trial. Twn•MA, Twn•MA#5 and Twn•MA#8 are variants of Towne obtained from Microbiological Associates. Twn•Aviron is highly-passaged Towne obtained at Aviron.

Figure 9. Schematic depiction of generation of chimeric CMV virus genomes by cotransfection of cosmids containing portions of Towne and Toledo genomes.

Figure 10. Schematic deptiction of the specific exemplary embodiments denoted Chimera I, Chimera II, Chimera III, Chimera IV, and Towne/Tol 11. Toledo genome is depicted as "Toledo"; highly-passaged Towne genome is depicted as "Towne•AV"; selected reading frames of importance, proposed function/homologues of selected ORFs, and scale (in kbp) is shown on the top line.

Figure 11. Replication of Toledo, highly passaged Towne, and Chimeras I, III, and IV (in order, respectively) in SCID-hu mice having a thymus/liver implant.

Figure 12. Schematic comparison of low-passage (long) Towne genome and high-passage (short) Towne genome.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage (Immunology - A Synthesis, 2nd Edition, E.S. Golub and D.R. Gren, Eds., Sinauer Associates, Sunderland, Massachusetts (1991)). Stereoisomers

(e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α,α-disubstituted amino acids, N-alkyl amino acids, lactic acid, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ε-N,N,N-trimethyllysine, ε-N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, ω-N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the righthand direction is the carboxy-terminal direction, in accordance with standard usage and convention. Similarly, unless specified otherwise, the lefthand end of single-stranded polynucleotide sequences is the 5' end; the lefthand direction of double-stranded polynucleotide sequences is referred to as the 5' direction. The direction of 5' to 3' addition of nascent RNA transcripts is referred to as the transcription direction; sequence regions on the DNA strand having the same sequence as the RNA and which are 5' to the 5' end of the RNA transcript are referred to as "upstream sequences"; sequence regions on the DNA strand having the same sequence as the RNA and which are 3' to the 3' end of the coding RNA transcript are referred to as "downstream sequences".

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence

is identical to a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of Fig. 1A-1R, or may comprise a complete cDNA or gene sequence. A full-length cDNA or gene sequence is defined as a polynucleotide containing the sequence(s) necessary to encode a complete protein product, including a translation initiation codon and a translation termination codon, unless linked to another encoding sequence in a format for production as a fusion protein. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window

may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of

the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence, for example, as a segment of an open reading frame shown in Fig. 1A-1R.

As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 90 percent sequence identity, more preferably at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). Preferably, residue positions which are not identical differ by conservative amino acid substitutions.

Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

The term "analog", "mutein" or "mutant" as used herein refers to polypeptides which are comprised of a segment of at least 10 amino acids that has substantial identity to a portion of the naturally occurring protein

The term "cognate" as used herein refers to a gene sequence that is evolutionarily and functionally related between species. For example but not limitation, in the human genome, the human CD4 gene is the cognate gene to the mouse CD4 gene, since the sequences and structures of these two genes indicate that they are highly homologous and both genes encode a protein which functions in signaling T cell activation through MHC class

II-restricted antigen recognition.

The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, an array of spatially localized compounds (e.g., a VLSIPS peptide array, polynucleotide array, and/or combinatorial small molecule array), a biological macromolecule, a bacteriophage peptide display library, a bacteriophage antibody (e.g., scFv) display library, a polysome peptide display library, or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues. Agents are evaluated for potential activity as antineoplastics, anti-inflammatories, or apoptosis modulators by inclusion in screening assays described hereinbelow. Agents are evaluated for potential activity as specific protein interaction inhibitors (i.e., an agent which selectively inhibits a binding interaction between two predetermined polypeptides but which does not substantially interfere with cell viability) by inclusion in screening assays.

As used herein, the terms "label" or "labeled" refers to incorporation of a detectable marker, e.g., by incorporation of a radiolabeled amino acid or attachment to a polypeptide of biotinyl moieties that can be detected by marked avidin (e.g., streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods). Various methods of labeling polypeptides and glycoproteins are known in the art and may be used. Examples of labels for polypeptides include, but are not limited to, the following: radioisotopes (e.g., ^{3}H, ^{14}C, ^{35}S, ^{125}I, ^{131}I), fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), enzymatic labels (e.g., horseradish peroxidase, β-galactosidase, luciferase, alkaline phosphatase), biotinyl groups, predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, transcriptional activator polypeptide, metal binding domains, epitope tags). In some embodiments, labels are attached by spacer arms of various lengths to reduce potential steric hindrance.

As used herein, "substantially pure" means an object

species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual macromolecular species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 to 90 percent of all macromolecular species present in the composition. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

The term "primer" as used herein refers to an oligonucleotide whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The primer is preferably single-stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with template. In some embodiments, the primers can be large polynucleotides, such as

from about 200 nucleotides to several kilobases or more. The primers herein are selected to be substantially complementary to the different strands of each specific sequence to be amplified. The primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, noncomplementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to be amplified to hybridize therewith and thereby form a template for synthesis of the extension product of the other primer.

The term "recombinant" used herein refers to macromolecules produced by recombinant DNA techniques wherein the gene coding for a polypeptide is cloned by known recombinant DNA technology. For example, an amplified or assembled product polynucleotide may be inserted into a suitable DNA vector, such as a bacterial plasmid, and the plasmid used to transform a suitable host. The gene is then expressed in the host to produce the recombinant protein. The transformed host may be prokaryotic or eukaryotic, including mammalian, yeast, Aspergillus and insect cells. One preferred embodiment employs bacterial cells as the host. Alternatively, the product polynucleotide may serve a non-coding function (e.g., promoter, origin of replication, ribosome-binding site, etc.).

DETAILED DESCRIPTION

Commonly-assigned U.S. patent application U.S.S.N. 08/414,926 filed 31 March 1995 is incorporated herein by reference.

The nomenclature used hereafter and the laboratory procedures in cell culture, molecular genetics, and nucleic acid chemistry and hybridization described below may involve well known and commonly employed procedures in the art. Standard

techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, and microbial culture and transformation (e.g., electroporation, lipofection). The techniques and procedures are generally performed according to conventional methods in the art and various general references (see, generally, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Oligonucleotides can be synthesized on an Applied Bio Systems oligonucleotide synthesizer according to specifications provided by the manufacturer.

Methods for PCR amplification are described in the art (PCR Technology: Principles and Applications for DNA Amplification ed. HA Erlich, Stockton Press, New York, NY (1989); PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfland, Snisky, and White, Academic Press, San Diego, CA (1990); Mattila et al. (1991) Nucleic Acids Res. 19: 4967; Eckert, K.A. and Kunkel, T.A. (1991) PCR Methods and Applications 1: 17; and U.S. Patent Nos. 4,683,202 and 4,965,188, each of which are incorporated herein by reference) and exemplified hereinbelow.

It is evident that optimal PCR and hybridization conditions will vary depending upon the sequence composition and length(s) of the targeting polynucleotide(s) and target(s), and the experimental method selected by the practitioner. Various guidelines may be used to select appropriate primer sequences and hybridization conditions (see, Maniatis et al., Molecular Cloning: A Laboratory Manual (1989), 2nd Ed., Cold Spring Harbor, N.Y.; Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA; PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfland, Snisky, and White, Academic Press, San Diego, CA (1990); Benton WD and Davis RW (1977) Science 196: 180; Goodspeed et al. (1989) Gene 76: 1; Dunn et al. (1989) J. Biol. Chem. 264: 13057 which are incorporated herein by reference.

A basis of the invention is the unexpected discovery

that there are significant genomic differences between clinical isolates of CMV and highly-passaged CMV strains, including differences between low-passage Towne and high-passage Towne, as well as differences as compared to Toledo strain; the identification of these genomic differences, including definition of novel genomic region(s); and the phenotypic significance and biological function of said genomic differences and specific ORFs within said novel genomic regions. Based, in part, on these unexpected discoveries, it is possible to construct and use chimeric CMV viruses which have predetermined genome compositions comprising at least a portion of a genome of a first CMV isolate or strain and at least a portion of a genome of a second (or subsequent) CMV isolate or strain, so as to form a complete, replicable recombinant chimeric CMV genome, with and the resultant chimeric CMV genome being capable of replication in a suitable host replication system and being useful for a variety of uses, such as human or veterinary vaccines, commercial reagents for laboratory use (e.g., as restriction enzymes are sold), use in screening systems to identify novel candidate drugs to inhibit replication or pathogenesis (e.g., virulence, tropism, host range, etc.) of pathogenic, clinically relevant CMV virus types, and other uses such as diagnostic reagents, gene expression vectors, anti-tumor agents, heterologous gene expression systems, and the like.

Overview

An approach of the invention starts with identification of DNA sequences which confer virulence on human cytomegalovirus (HCMV). These sequences can be manipulated to produce a new, more efficacious HCMV vaccine strain with predicted characteristics. Introduction of the virulence genes into an overattenuated strain can improve its immunogenicity and deletion of the virulence genes from a virulent strain can render it safe in humans by decreasing its virulence. Specifically, deletion of genetic information from a clinical isolate called Toledo is used to attenuate an HCMV virus, and in one embodiment, a segment from a laboratory strain called Towne, especially a highly-

passaged Towne variant, is transferred to the deleted region of Toledo to act as a "spacer". Deleting genetic information has utility in improving a clinical isolate such as Toledo as an immunizing composition. Removing these sequences from Toledo, which has been shown to cause disease in people, can result in an attenuated virus which may be a safe vaccine candidate.

The Towne strain of HCMV has been used as a vaccine in humans. In some clinical settings, Towne has been used to prevent the disease consequences associated with infection by HCMV (reviewed by Marshall and Plotkin In: *The Human Herpesviruses* B. Roizman, R.J. Whitley, & C. Lopez Eds. Raven Press, New York). The Towne strain is believed to be overattenuated as a vaccine candidate and consequently, is poorly immunogenic. This loss of immunogenicity may have been the result of an extensive passage history in tissue culture. Genetic information in the virulence region may have been lost during passage, particularly after about Passage 40. Variation in DNA content among isolated strains does exist based on crude hybridization experiments. Other investigators have reported minor regions of sequence heterogeneity between two so-called laboratory strains of HCMV, the Towne and AD169 strains. Heterogeneities can exist within HCMV strains depending upon the extent of passages in their culture history.

The public health impact of HCMV infections have not been well controlled by current treatment strategies or available antiviral chemotherapies. Preventive vaccine strategies are likely to prove efficacious because of the observations that seropositive renal allograft recipients are protected from severe HCMV disease and maternal immunity protects the fetus from disease after intrauterine infection. HCMV (Towne) was developed as a vaccine strain by serial passage 125 times in WI38 human diploid fibroblasts (Towne 125). It has been administered to humans without significant adverse reactions. However, in one study, vaccinees were directly challenged by wild-type virus and found to resist only low challenge doses of 10 plaque-forming units or less. The consensus view is that the Towne strain may be overly attenuated. One positive feature of the Towne strain

is that it has never been shown to reactivate.

One important obstacle to the development of a vaccine for HCMV is the lack of an animal model system that can be used to test the safety and efficacy of vaccine candidates. Therefore, cell culture systems or surrogate animal models such as the SCID-hu (thy/liv) mouse have to be developed to test vaccine strains. Replicative differences in HCMV strains have been described in a variety of cell types and in the SCID-hu mouse model. These differences correlate to the virulence and passage history of the virus. Thus, low passage, virulent clinical isolates, such as Toledo, can replicate better in the human implant of SCID-hu (thy/liv) mice and in cultures of human endothelial cells than cell culture adapted, highly-passaged avirulent laboratory strains such as Towne or AD169 (Brown *et al.* 1995; Waldman *et al.*, 1991). This observation can be exploited to measure the "virulence" of a strain by assessing its growth characteristics in the SCID-hu mouse, in vivo in humans, or by other means. Recombinant vaccine candidates such as the ones described here which have deleted or incorporated DNA sequences are believed to replicate less well than the virulent parent in a suitable virulence assay. This observation would be indicative of an attenuated vaccine candidate. Deletion of the Toledo UL/b' region from the low passage, virulent HCMV Toledo genome results in a virus with reduced replicative ability in the SCID-hu mouse. This recombinant virus should have a concomitantly reduced virulence which allows administration of the virus without causing the undesired clinical manifestations exhibited by the Toledo virus in humans.

The invention identifies, maps, and sequences differences between the virulent Toledo strain and the avirulent highly passaged Towne strain, for the purpose of transferring novel genetic information to Towne to restore its immunogenicity or, alternatively, to remove information from Toledo to render it safe as a vaccine candidate. One major region of difference mapped to the internal portion of the L component. This large 13kbp region present in Toledo but not highly passaged Towne is located at the border between the unique long (UL) and the

inverted repeats bordering the UL region termed IRL or b'. We have deduced the coding information resident in the Toledo sequences and have extensively compared the information resident in AD169, highly passaged Towne and Toledo. We have made recombinant viruses which have either inserted the UL/b' region from the virulent Toledo strain, into the corresponding region of Towne, and have also deleted this region from Toledo and replaced it with a selectable marker and reporter gene or with the corresponding UL/b' region from Towne. Deletion of the virulence genes from Toledo decreased the ability of the recombinant to replicate within the SCID-hu (thy/liv) mouse, a model for CMV virulence. The new recombinant viruses exhibit growth properties in the SCID-hu mouse that indicate that vaccine candidates with attenuated virulence can be generated by deleting the UL/b' region from the Toledo virus. We have also demonstrated that we can add the Toledo region to the Towne virus which will presumably result in increased immunogenicity for the highly passaged Towne virus while retaining its safe profile for humans.

Figures 1A-1R show the nucleotide sequence of Toledo genome region isolated from Toledo strain of HCMV. Figures 2A-2H show the deduced amino acid sequences of open reading frames UL130 through UL151.

A basis of the present invention is the surprising and unexpected finding that: (1) clinical isolates of pathogenic CMV variants contain a genomic region which typically is not present in CMV strains which have undergone extensive laboratory passaging of the virus in cell culture, and (2) functional disruption (e.g., deletion or insertional inactivation and the like) of genes in this genomic region produces a substantial attenuation of CMV virulence and pathogenicity in vivo. The genomic region is conveniently termed the "Toledo genomic region" herein, although equivalent (e.g., homologous) regions or subsequences thereof are present in other clinical isolates of CMV besides the Toledo strain of CMV; the term "Toledo genomic region" encompasses these homologous regions in other clinical CMV isolates and non-isolated pathogenic CMV variants which have a genomic region of at least 500 bp having at least 80 percent

sequence identity to the Toledo genomic region of the Toledo strain having the sequences disclosed herein and in WO96/30387, incorporated herein by reference. The Toledo genomic region which is present in pathogenic CMV isolates and which is typically substantially absent in laboratory passaged CMV strains (e.g., AD169, Towne) has been sequenced and several open-reading frames have been identified. Functional disruption of these open reading frames, either singly or in combination, has been unexpectedly found to substantially reduce virulence of the resultant CMV mutant(s) in vivo. Thus, in part, the invention provides methods and compositions for suppressing or inactivating expression of genes of the Toledo genomic region and its homolog regions in other CMV variants, and thereby reducing virulence and pathogenicity of clinically important CMV variants. The invention is, in part, further based on the heretofore unrecognized finding that pathogenic clinical isolates of CMV have a distinct genome as compared to the commonly used laboratory-passaged strains of human CMV (e.g., AD169, Towne), and that the genomic region which is present in the clinical isolates and which is substantially absent in laboratory-passaged strains confers enhanced virulence in vivo. Most common approaches to development of CMV therapies and vaccines have heretofore relied on laboratory-passaged strains which lack the Toledo genomic region and the genes encoded therein which have been unexpectedly found to confer enhanced in vivo virulence and are believed to contribute to clinical pathology and CMV-related disease.

The invention provides a method for attenuating virulence of CMV comprising functionally inactivating at least one open reading frame in a genomic region of a CMV genome having substantial sequence identity to at least 300 bp, typically at least 500 bp, of an approximately 15 kb sequence present in the genome of the Toledo strain of CMV and absent from the genome of the AD169 strain of CMV. In an aspect, the method functionally inactivates at least one open reading frame present in a genomic region of a CMV genome having substantial identity to at least 300 bp of a 13 kb sequence present in the genome of the Toledo

strain of CMV and absent from the genome of the highly -passaged Towne strain of CMV. In an embodiment, the method functionally inactivates at least one open reading frame present in a genomic region of a CMV genome having substantial identity to at least 500 bp of the sequence shown in Figs. 1A through 1T. In an embodiment, the method functionally inactivates at least the open reading frame corresponding to UL148 as identified herein. In a variation, the method funtionally inactivates open reading frames in the region spanning UL138 to UL148. In an embodiment, the method functionally inactivates UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148. In a variation, UL148 is inactivated singly or in combination with other open reading frames of the Toledo genomic region. In a specific embodiment, UL148 is inactivated in combination with UL141 and/or UL144. Inactivation is typically accomplished by genetic engineering and involves predetermined mutations (which may include additions, transpositions, or deletions), generally of the specific type which are not known to occur naturally in CMV strains even after extensive passaging.

In an aspect, the method of attenuating virulence comprises functional inactivation of open reading frames by structural mutation (e.g., deletion, insertion, missense or nonsense mutation, and the like) of at least one open reading frame, or a mutation of a transcriptional control sequence that controls transcription of the open reading frame, or mutation of a splicing signal sequence or the like necessary for efficient expression of the encoded gene product of the open reading frame. In an embodiment, a selectable marker gene is introduced into an open reading frame, often in the portion of the open reading frame believed to encode the amino-terminal two-thirds of the gene product, to structurally disrupt the open reading frame and result in the inactivation of the open reading frame's capacity to encode its functional gene product. In a variation, open reading frame UL148 is structurally disrupted by predetermined mutation, often produced by site-directed mutagenesis or in vitro recombination; in one embodiment the structural disruption results from insertion of a selectable and/or screenable marker

gene (e.g., *gpt/lacZ*). In an embodiment, a selectable marker gene is used to replace all or part of at least one open reading frame, such as by replacement of a deleted region of the Toledo genomic region with a selectable marker gene. In a variation, a region spanning open reading frame UL138 to UL148 is structurally disrupted by predetermined mutation; in one embodiment the structural disruption results from deletion of the UL138-UL148 region and replacement with a selectable and/or screenable marker gene (e.g., *gpt/lacZ*).

In an aspect, the functional inactivation of a Toledo genomic region gene is provided by transcriptional and/or translational suppression with an antisense polynucleotide having a sequence of at least 15 nucleotides, typically at least 25 nucleotides, that are substantially complementary to a Toledo genomic region, most usually the antisense polynucleotide is substantially complementary to an open reading frame sequence of a Toledo genomic region open reading frame. In an embodiment, the antisense polynucleotide is substantially complementary to at least 25 nucleotides of UL148. In an embodiment, the antisense polynucleotide is complementary to UL148 and further comprises additional 5' and/or 3' nucleotide(s) which are not substantially complementary to UL148. In variations, the antisense polynucleotides comprise non-natural chemical modifications, and can include, for instance, methylphosphonates, phosphorothioates, phosphoramidites, phosphorodithioates, phosphorotriesters, and boranophosphates. In a variation the antisense molecules can comprise non-phosphodiester polynucleotide analogs wherein the phosphodiester backbone is replaced by a structural mimic linkage include: alkanes, ethers, thioethers, amines, ketones, formacetals, thioformacetals, amides, carbamates, ureas, hydroxylamines, sulfamates, sulfamides, sulfones, and glycinylamides. In a variation, the invention provides peptide nucleic acids (PNAs) having a nucleobase sequence which is substantially complementary to a Toledo genomic region sequence, such as an open reading frame (e.g., UL148, UL141, UL142, etc.).

The invention also provides attenuated live virus CMV

vaccines wherein at least one open reading frame of a Toledo genomic region is structurally disrupted by predetermined mutation. Typically, the UL148 open reading frame is structurally disrupted, either singly or in combination with other Toledo region open reading frames (e.g., UL141, UL144, and the like). Often the disruption of the open reading frame is an insertion, deletion, or replacement mutation which confers the property of reduced virulence as determined by a suitable in vivo virulence assay (e.g., see Experimental Examples). Toledo genomic region mutants which exhibit at least one log reduction, preferably two logs or more reduction, in virulence as determined by in vivo virulence assay, or other equivalent virulence measure, are attenuated CMV vaccines. Such attenuated CMV vaccines are used to immunize individuals to confer protective immunity, typically antibody-mediated and/or cell-mediated immunity, to prevent or reduce the severity of subsequent CMV infection following a suitable immunization period.

In an aspect, the invention also provides attenuated live virus CMV vaccines wherein at least one open reading frame of a Toledo genomic region is replaced by a segment of the Towne genome which is not present in AD169. The highly-passaged Towne genome comprises a region not present in AD169; the region contains open reading frames designated UL147, UL152, UL 153, and UL154 and generally is spanned by nucleotides 178221 to 180029 of the Towne genome according to the AD169 numbering convention. An attenuated virus of the invention can, in one embodiment, comprise a Toledo genome wherein the Toledo genome region spanning open reading frames UL133 to UL151 are replaced with a Towne genome region spanning UL147, UL152, UL153, and UL154; this engineered CMV virus variant is an attenuated Toledo virus which comprises desirable features of Towne while reducing undesirable virulence of the Toledo genome region. The invention provides other variations of this basic method, whereby a segment of the Toledo genome region comprising at least one open reading frame is deleted or otherwise structurally disrupted in a CMV variant having a Toledo genome region or its homolog, and a segment of a Towne genome region comprising at least one open reading frame

in inserted in the CMV variant. In an embodiment, the engineered CMV variant comprises: (1) Toledo DNA (DNA substantially identical to a Toledo strain, preferably identical to it) from about nucleotides 1 to about 168,000 corresponding to (i.e., according to) the AD169 nucleotide numbering convention, operably linked to (2) Towne DNA (DNA substantially identical to a Towne strain, preferably identical to it) from about nucleotides 143,824 to 189,466 according to the AD169 nucleotide numbering convention, operably linked to (3) Toledo DNA (DNA substantially identical to a Toledo strain, preferably identical to it) from about nucleotides 189,466 to about 209,514 corresponding to (i.e., according to) the AD169 nucleotide numbering convention, operably linked to (4) Towne DNA (DNA substantially identical to a Towne strain, preferably identical to it) from about nucleotides 200,080 to 229,354 according to the AD169 nucleotide numbering convention. The invention also provides vaccine compositions and formulations of such attenuated CMV viruses, which can include adjuvants, delivery vehicles, liposomal formulations, and the like. The invention also provides the use of such attentuated CMV variants for prevention of CMV disease and infection; in one aspect this use includes administration of such vaccine to human subjects.

In a variation, the functional inactivation of a Toledo genomic region gene is provided by suppressing function of a gene product encoded by a Toledo region open reading frame by contacting or administering an antibody which is specifically reactive with said gene product. In an embodiment, the Toledo genomic region gene is UL148, UL141, and/or UL144, typically at least UL148, although other Toledo open reading frames can be used. The antibody binds to a gene product encoded by a Toledo region open reading frame with an affinity of at least about $1x10^7$ M^{-1}, typically at least about $1x10^8$ M^{-1}, frequently at least $1x10^9$ M^{-1} to $1x10^{10}$ M^{-1} or more. In some aspects, the antibody is substantially monospecific. In an embodiment, the antibody is a human antibody raised by immunizing an individual with an immunogenic dose of a gene product of a Toledo region open reading frame. In an embodiment, the human antibody is a

monoclonal antibody, or collection of human monoclonal antibodies which bind to the Toledo region gene product(s). In an embodiment, the antibody is a humanized antibody comprising complementarity-determining regions substantially obtained from a non-human species immunoglobulin reactive with the Toledo region gene product, and further comprising substantially human sequence framework and constant regions. The invention also comprises pharmaceutical formulations of such antibodies and the use of such antibodies to treat or prevent CMV diseases, such as by passive immunization or the like.

In an aspect, the invention provides a composite CMV variant comprising a Towne genome and at least one open reading frame of a Toledo genome region, typically present in or adjacent to the U_L/b' region of the composite CMV. In an aspect, the composite CMV is a Towne genome further comprising a Toledo UL148, UL141, and/or UL144. In an embodiment, the composite CMV is a highly-passaged Towne genome with a complete Toledo genome region; in a variation said Toledo genome region has at least one open reading frame functionally inactivated to further attenuate the virulence of the composite CMV.

In a variation, the invention provides a diagnostic method for identifying a virulent CMV strain in a sample by detecting the presence of unique Toledo genome region polynucleotide sequences and/or by detecting the presence of a polypeptide encoded by an open reading frame of the Toledo genomic region. Detection of polynucleotide sequences can be by any suitable method, including but not limited to PCR amplification using suitable primers, LCR, hybridization of a labeled polynucelotide probe, and the like. Detection of polypeptide speceis is typically done by immunoassay using a pecific antibody to the Toledo region gene product(s).

The invention also provides a method of treating or preventing CMV infection, the method comprising administering to an individual an efficacious dose of a polypeptide which is substantially identical to the deduced amino acid sequence of UL148. In a variation, the polypeptide is a truncated variant, mutein, or analog of the deduced amino acid sequence of UL148,

wherein the polypeptide is soluble.

5

EXPERIMENTAL EXAMPLES

Overview

The growth advantage of Toledo in the SCID-hu mouse model resides in the genetic information encoded by the additional sequences (Toledo genomic region) we have identified. One gene in particular, UL148, has been mutagenized in Toledo by insertion of a selectable marker *(gptILacZ)* and the Toledo-based recombinant has been shown to replicate less well than Toledo in the SCID-hu assay. The genetic information of the corresponding region of the avirulent Towne virus has been deduced by nucleotide sequence analysis and demonstrated to lack an open reading frame in Towne. UL148 can be considered to be representative of a "virulence determinant" for Toledo. The new Toledo sequence identified at the inverted repeats has been analyzed to reveal novel genes in Toledo. Deletion of genes encompassing UL138 to UL148 in recombinant viruses have been tested for growth properties in the SCID-hu (thy/liv) mouse. These recombinants have been shown to replicate to levels similar to the Towne virus and represent attenuated vaccine candidates, since Towne has been shown to be safe and avirulent in humans. Such recombinants should show increased immunogenicity owing to their greater similarity to low passage virulent strains over that shown by highly-passaged Towne in humans. In addition, these strains should not exhibit the fully virulent phenotype shown by unmodified Toledo in humans due to the alterations we have introduced into their genomes.

This invention describes new recombinant HCMV viruses not previously described which are attenuated in virulence relative to low passage, virulent isolates by virtue of deletion of sequences shown to be present in low passage, virulent isolates but which are lacking in laboratory strains. The identification of these sequences was essential in order to prepare transfer vectors capable of shuttling deletions (or insertions such as selectable markers) resulting in an effective removal of coding information. Knowledge of the ORF usage on these DNAs permits deletion or insertion of one DNA into the other to specifically disrupt existing coding information. In

addition, this invention identifies sequences which can be used as **"spacer"** DNA for substitution into deleted regions of HCMV clinical isolates for purposes of attenuation.

Cosmid subclones of Towne and Toledo

Cosmid subclones of the CMV(Towne) and CMV(Toledo) genomes were constructed according to the method of Kemble et al. (1996) J. Virol. 70: 2044, incorporated herein by reference. Human foreskin fibroblast (HF) cells were infected with either Towne or Toledo and following the development of extensive CPE, DNA was isolated from nucleocapsids by a procedure similar to that used for the preparation of HSV nucleocapsids (Denniston et al. (1981) Gene 15: 365, incorporated herein by reference). The DNA was partially digested with *Sau3*AI, fractionated by agarose gel electrophoresis, and ligated to the *Bam*HI site of *Bam*HI, *Xba*I digested arms of the SuperCos•A1 cosmid vector. SuperCos•A1 was derived from SuperCos-1 (Stratagene, San Diego, CA) by the insertion of an oligonucleotide incorporating *Srf*I and *Pac*I recognition sequences flanking a unique *Bam*HI site. The position of the cosmid subclones relative to the viral genome was identified by Southern and DNA sequence analyses.

Overlapping cosmids for virus regeneration

Mapping the extent of the viral insert within the cosmid subclones was used as a basis to form specific Towne/Toledo chimeric viruses by choosing the appropriate cosmids from each virus. The ends of adjacent cosmids should overlap (~ 200bp or more) such that homologous recombination is permitted in eucaryotic cells.

To construct a Toledo based virus which lacked the Toledo U_L/b' region and in its place contained the Towne U_L/b' region, the following set of cosmids was used: Tol29, Tol58, Tol182, Tol22, Tol158, Tol24, Tn39, and Tn50. The resulting virus was designated Tol/Twn 39/50 (see Fig. 5). Other viruses were regenerated by cosmid cotransfection which lacked portions of the Toledo U_L/b' region. Toledo based viruses were generated by the cotransfection of the Toledo cosmids, Tol29, Tol58, Tol182, Tol22, Tol158, Tol24, Tol 212, Tol187 OR Tol59, Tol150,

Tol239, Tol235, Tol158, Tol24, Tol212, Tol187. Towne/Toledo chimeras lacking portions of the Toledo U_L/b' region were regenerated by cotransfection of Tn43, Tn13, Tn24, Tn9, Tn42, Tn51, Tol212, Tol187. Because Tol 212 and Tol187 did not overlap, deletions resulted in the viruses regenerated from these cosmid sets which lacked varying portions of the Toledo U_L/b' region.

Preparation of cosmids for cotransfection

A set of overlapping cosmid clones constituting the appropriate viral genome were individually digested with *Pac*I to release the intact viral insert from the cosmid vector. The restriction enzyme was inactivated by heating at 65°C for 20 minutes, the cosmids were combined and the DNA precipitated with ethanol. A $CaPO_4$ precipitate was formed from approximately 8 to 16μg of this mixture and transfected using general transfection methods. The DNA was transfected into approximately 1 X 10^6 low passage (<15 passes) HF, LF (human embryonic lung fibroblast) or IFIE1.3 (a gift of Ed Mocarski; these cells are immortalized HF cells that express the CMV major immediate early protein) cells. All these cells are permissive for CMV replication.

For HF and LF cells, approximately 1 X 10^6 cells were plated onto a 25cm^2 flask 3 to 5 hours prior to the addition of the DNA-$CaPO_4$ precipitate. At this point, the precipitate was adsorbed directly to the cell monolayer for 30 minutes prior to the addition of media. 2ml of media was added and incubation continued for 4 hours at 37°C.

For IFIE1.3 cells, the cells were trypsinized approximately 16 hours prior to the addition of the DNA-$CaPO_4$ precipitate and seeded at a 1:2 density. At the appropriate time post seeding, the DNA-$CaPO_4$ precipitate was added in addition to 2ml of media and incubated at 37°C for 4 hours.

Following the 4 hour incubation, the DNA-$CaPO_4$ precipitate was removed, the cells incubated at 37°C for 3 min in 15% glycerol in Hepes buffered saline, rinsed one time with media and fed with 5ml of media. The media on the cells was changed every 3 to 4 days and plaques appeared in 10 to 21 days.

Construction of recombinant CMV by insertion of a *gpt/LacZ* marker

Two plasmids encompassing the Toledo U_L/b' region and derivatives thereof were constructed which contained a marker gene. A segment of DNA encompassing AD169 bases 156251-174483 was removed from pON2601 (Cha et al. (1996) J. Virol. 70: 78, incorporated herein by reference) and a *Pac*I linker was introduced at AD169 base 174484 to yield a subclone of pON2601. Figs. 3 and 4 show a schematic drawing of the open reading frames in the Toledo U_L/b' region using sequence numbering from the Toledo U_L/b' region DNA insert. A 4.8 kb DNA fragment containing the E. coli *gpt* and *lac Z* genes driven by the HSV thymindine kinase and β actin promoters (Prichard et al. (1996) J. Virol. 70: 3018, incorporated herein by reference), respectively, was then inserted into the *Nsi*I site in Toledo UL148 within the pON2601 subclone. The resulting plasmid containing the *gpt* and *lacZ* insert in UL148 was designated pGD6. Toledo open reading frames UL138 to UL148 were removed from pGD6 by a *Bam*HI collapse to produce the plasmid pGD7. Toledo recombinant viruses Tol pGD6 and Tol pGD7 were constructed using plasmids pGD6 and pGD7, respectively, as described (Prichard et al. (1996) op.cit.

Analysis of recombinant CMV in SCID-hu (thy/liv) mice

SCID-hu (thy/liv) mice were derived by implanting human fetal thymus and liver beneath the kidney capsule of a female C.B. -17 *scid/scid* IcrTac mouse (McCune et al. (1988) Science 241: 1632, incorporated herein by reference). The SCID-hu (thy/liv) mouse model serves as an animal model that can distinguish virulent from avirulent strains of CMV based on their replication levels within the human implant (Mocarski et al. (1993) Proc. Natl. Acad. Sci. (U.S.A.) 90: 104 and Brown et al. (1995) J. Infect. Dis. 171 1599, each incorporated herein by reference). Several weeks following implantation, the human implant on the murine kidney was surgically exposed and an inoculum of ~10^4 PFU of the appropriate virus was injected directly into the human tissue in a volume of 10 - 25μl. The murine kidney/human implant was placed back into the animal in

its natural position and the animal was recovered. 2 weeks following infection of the human tissue, the animal was sacrificed and the implant was removed and added to 2ml of 4.5% skim milk/50% media.

The excised implant was homogenized with an automated Dounce apparatus (Glas-Col, Terre Haute, IN) and the suspension was stored at -80°C until the titers were determined. The suspension was thawed at 37°C, sonicated on ice by three cycles of 10 sec on/ 10 sec off and centrifuged to remove the debris. The supernatent was recovered and the titer of CMV present was determined on confluent monolayers of HF cells. 7 to 10 days after plating the virus, the monolayers were fixed and stained with Giemsa and plaques enumerated.

Fig. 6 shows results from this experiment. The virulence of the Toledo strain CMV is attenuated by functional disruption of Toledo genome region open reading frames.

The difference in virulence between the Towne and Toledo strainsappears to have resulted from genetic differences generated during the adaptation of Towne to growth in dipoid fibroblasts in culture. Both Towne and Toledo were originally isolated from the urine of a congenitally infected infant. Towne was subsequently passaged over 125 times in culture resulting in genetic alterations in the viral genome and an avirulent virus. The virulent Toledo virus, in contrast, was passaged approximately 5 times in diploid fibroblasts in order to produce material that could cause disease in humans.

These linked genetic and biological differences can be used to create a live, attenuated HCMV vaccine. The rationale for tissue culture adaptation of Towne was to generate a live, attenuated vaccine strain. Towne has been shown to be safe and somewhat immunogenic. Towne, however, is overattenuated. The immune response induced by inoculation with Towne does not protect against subsequent HCMV infection as effectively as that generated by natural infection. Vaccine candidates can be generated by replacing genetic elements of the overattenuated Towne strain with homologous portions of the virulent Toledo strain. Through the analysis of these "chimeric" viruses, a

skilled artisan can select those that have the level of desirable characteristics of Towne and be attenuated to a more efficacious degree.

Our first four chimeric viruses, as a set, will contain the entire Toledo genome introduced into the Towne genetic background. Each individual chimera of the set will contain approximately 40-55% of the Toledo genome; the remainder will be derived from Towne. Each of these chimeras will contain the UL/b' region derived from Toledo. Genes within this region of the Toledo genome can affect cell tropism of HCMV. The viruses, designated chimera I, II, III, and IV were constructed from the cosmids listed in Table 1 (see also Figs. 9 and 10).

Table 1. Cosmids used to generate specific chimeras.

Viruses	I	II	III	IV
	Tn46a	Tn46	Tn46	Tol29
	Tol58b	Tn45	Tn45	Tol58
	Tol182	Tol239	Tn23	Tn23
	Tn47	Tol22	Tn47	Tn47
	Tn44	Tol158	Tol184	Tn44
	Tn26	Tn26	Tol24	Tn26
	Tn20	Tn20	Tol212	Tol212
	Tol11	Tol11	Tol11	Tol122

Large quantities of each cosmid were prepared by purification of the E.coli produced material over a Qiagen column as described by the manufacturer. 10 micrograms of each cosmid was digested with the restriction enzyme Pac I (New England Biolabs) to physically separate cosmid vector from viral sequences. Following digestion, the enzyme was inactivated by incubation at 65oC for 20 minutes and the appropriate cosmids were combined, precipitated with ethanol in the presence of 0.3M sodium acetate, rinsed in 70% ethanol, and air-dried briefly. The resulting DNA was solubilized in approximately 100

microliters of 10mM Tris pH 7.5/1mM EDTA. Various amounts of the cosmid mix was transfected by the calcium phosphate technique into permissive fibroblast cells, specifically human lung fibroblasts (LF, prepared in our laboratory), human neonatal foreskin fibroblast (HF, a gift of Dr. Ed Mocarski, Stanford University), MRC-5 (ATCC) or IFIE1.3 cells (a gift of Ed Mocarski, Stanford University). The IFIE1.3 cell line constitutively expresses the HCMV ie1 gene product and has been transformed with the human papilloma virus E6 and E7 genes transduced by a retrovirus vector. 3 to 5 hours after transfection the cells were shocked by incubation in 15% glycerol/Hepes buffered saline for 3 minutes at 37oC and fed with DME/10% fetal bovine serum. 7 to 10 days after transfection plaques with distinct HCMV CPE were evident.

Plaques derived from the chimeras were allowed to grow until 100% of the monolayer exhibited CPE. At this point, DNA was extracted from the supernatent and cellular fractions and analyzed by restriction enzyme digestion. The structures of the viruses can be deduced by the cosmids used for construction of the chimera and confirmed by comparing the EcoRI digestion pattern to the maps derived for Towne and Toledo (see Fig. 10). Table 2 describes the composition of each of the chimeras, the nucleotide limits are derived from sequence analysis of the end of each cosmid insert and its homology to the AD169 strain of HCMV, which has been sequenced in its entirety. All of the chimeras had restriction enzyme patterns consistent with the proposed structures.

Table 2. Genetic composition of the chimeras.

Chimera	Towne DNA	Toledo DNA	Crossover Region
I	1 - 3799	15750-67568	3800-15749
	81647-170499	175069-203136	67569-81646
	205803 to S term.		170500-175068

II	1 - 47985 138834 - 170499 205803 to S term.	53244 --~110000 175069 - 203136	47986-53243 ~110000- 130833 170500 - 175068 203137-205802
III	1 - ~99000* 205803 to S term	108094 - 203136	~99000-108093 203137-205802
IV	43981 - 145583	1 - 41356 150754 - S term	41357 - 43980 145584 - 150753

* Sequence at end of cosmid was undefinable. Nucleotide number is not exact crossover region is the region of cosmid overlap. The contribution ofeach virus to this region has yet to be defined.

Two other chimeras are constructed based on an observation derived from the sequence analysis of several different members of the beta-herpesvirus family including HCMV, human herpesvirus 6 (the causative agent of roseola) and murine cytomegalovirus. Representative members of each of these viruses have been sequenced in their entirety and a "core" set of genes corresponding to HCMV UL23 to UL122 are conserved among these evolutionarily divergent entities (Chee, et al. 1990; Gompels, et al. 1995; Rawlinson, et al. 1996, incorporated herein by reference). These core genes contribute to DNA replication, virion structure, and other basic features of the virus. Genes outside this core region are involved in virus-host and virus-immune system interaction and may determine specific properties of virus biology. Replication of the Chimeras was tested in SCID-hu mice having a thy/liv sandwich under the kidney capsule; representative data is shown in Figure 11.

Two additional chimeras are constructed: one which has the core derived from Toledo with the remainder of the genes derived

from Towne and an inverse construct in which the core is derived from Towne and the remainder of the genes are from Toledo. These viruses are constructed through the use of overlapping cosmids and derivatives of the cosmids. Table 3 outlines the constructs that can be used to generate these two chimeric viruses.

Table 3. Construction of chimeras containing conserved core regions.

Towne Core/Tol Noncore	Tol Core/Towne Noncore
Tol 29	Tn43
Tol58 nts: 3800 - 27862	Tn45 nts: 7854 - 27862
Tn45 nts: 27500 - 53243	Tol 58: 27500 - 43980
Tn23	Tol 182
Tn47	Tol 22
Tn44	Tol 158
Tn26	Tol 24
Tn20	Tol212 nts:145584-17200
Tol11nts: 170852 - 188890	Tn 39 nts:170852 - 183512
Tol122	Tn 15

All of these viruses are used to inoculate healthy adult human volunteers. These individuals are assessed for symptoms of HCMV disease, including fever, malaise, and abnormal liver enzyme levels. Hallmarks of viral infection are also assessed by measuring HCMV specific antibody titers before and after inoculation as well as viral culture for the isolation of infectious virus from bodily fluids. A successful vaccine candidate is identified as a strain that maintains the safety profile of Towne while stimulating a greater immune response to the virus.

Figure 12 shows schematic depiction of the Toledo genome incomparison with highly passaged Towne (short genome) and low-passage Towne (long genome).

The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and many modifications and variations are possible in light of the above teaching.

Such modifications and variations which may be apparent to a person skilled in the art are intended to be within the scope of this invention.

CLAIMS:

1. A method for attenuating a cytomegalovirus comprising functionally disrupting an open reading frame of a Toledo genome region or its homologs.

2. A method of claim 1 for attenuating a CMV strain or isolate containing an encoding polynucleotide sequence encoding a polypeptide which is at least 80 percent sequence identical to a polypeptide encoded by UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148 of the Toledo genomic region; the method comprising functionally inactivating said encoding polynucleotide sequence to produce a Toledo region-attenuated CMV variant.

3. A chimeric CMV virus, comprising a genome having a plurality of polynucleotide sequences, linked in conventional phosphodiester linkage, wherein at least two of said polynucleotide sequences are derived from different clinical isolates or strains of CMV.

4. The chimeric CMV virus of claim 3, wherein said chimeric CMV virus comprises a genome having a plurality of polynucleotide sequences, linked in conventional phosphodiester linkage, wherein a first CMV genome sequence of at least 500 bp and less than a complete CMV genome length is at least 98 percent sequence identical to a first CMV isolate or strain, and at least one additional CMV sequence of at least 500 bp and less than a complete CMV genome length is at least 98 percent sequence identical to a second CMV isolate or strain which has a genome having a polynucleotide sequence of at least 500 bp which is less than 60 percent sequence identical to any portion of the genome of said first CMV isolate or strain and/or which is absent or substantially absent in the genome of said first CMV isolate or strain.

5. The chimeric CMV virus of claim 4, wherein the chimeric

CMV virus genome is at least 80 percent sequence identical to a high-passage Towne genome and said chimeric CMV virus genome typically further comprises genetic information which is at least 80 percent sequence identical to at least 1 kbp of a virulence region of a clinical isolate of CMV or a low-passage strain of CMV other than low-passage Towne.

6. A chimeric CMV virus of claim 5, wherein a complete Toledo genome region is present.

7. A chimeric CMV virus of claim 3 which is Chimera I.

8. A chimeric CMV virus of claim 3 which is Chimera II.

9. A chimeric CMV virus of claim 3 which is Chimera III.

10. A chimeric CMV virus of claim 3 which is Chimera IV.

11. A chimeric CMV virus of claim 3 which is Towne/Tol11.

12. A method for producing an attenuated CMV virus, comprising forming a chimeric CMV virus having a genome comprising a first genome portion derived from a CMV clinical isolate or CMV strain having a virulence region and a second genome portion derived from a highly passaged CMV strain or isolate lacking a virulence region.

ABSTRACT

A method is provided for attenuating a cytomegalovirus comprising functionally disrupting an open reading frame of a Toledo genome region or its homolog and making chimeric CMV virus genomes.

Toledo UL134

```
        10         20         30         40         50         60
MARTREASPV PPRSPMPSHI HTMIFSPAWN LKLRVGKGRC TDIYALDFWK RHFLARNVFI

        70         80         90        100        110        120
VQTLRKEMCA KSENSLSHRG RVTFRSDAAA VVVEPRPRPP ARQLVPPRPR RVASAAWRGE

       130        140        150        160        170        180
ARRADRRALP SAATVVVNSP SVRTEVCLSV YPSVYLSPYL SSVWVPMSVL AAAVG*....
```

Toledo UL135

```
        10         20         30         40         50         60
MSVHRPFPTR SLRFQAGEKI MVWIWLGIGL LGGTGLASLV LAISLFTQRR GRKRSDETSS

        70         80         90        100        110        120
RGRLPGAASD KRGACACCYR NPKEDVVEPL DLELGLMRVD THPPTPQVPR CTSLYIGEDG

       130        140        150        160        170        180
LPIDKPEFPP ARFEIPDVST PGTPTSIGRS PSHCSSSSSL SSSTSVDTVL YQPPPSWKPP

       190        200        210        220        230        240
PPPGRKKRPP TPPVRAPTTR LSSHRPPTPI PAPRKNLSTP PTKKTPPPTK PKPVGWTPPV

       250        260        270        280        290        300
TPRPFPKTPT PQKPPRNPRL PRTVGLENLS KVGLSCPCPR PRTPTEPTTL PIVSVSELAP

       310        320        330        340        350        360
PPRWSDIEEL LEQAVQSVMK DAESMQMT*. .......... .......... ..........
```

Toledo UL136

```
        10         20         30         40         50         60
MSVKGVEMPE MTWDLDVRNK WRRRKALSRI HRFWECRLRV WWLSDAGVRE TDPPRPRRRP

        70         80         90        100        110        120
TWMTAVFHVI CAVLLTLMIM AIGALIAYLR YYHQDSWRDM LHDLFCGCHY PEKCRRHHER

       130        140        150        160        170        180
QRRRRQAMDV PDPELGDPAR RPLNGAMYYG SGCRFDTVEM VDETRPAPPA LSSPETGDDS

       190        200        210        220        230        240
NDDAVAGGGA GGVTSPATRT TSPNALLPEW MDAVHVAVQA AVQATVQVSG PRENAVSPAT

       250        260        270        280        290        300
*.......... .......... .......... .......... .......... ..........
```

Fig. 2B

Toledo UL130

```
        10         20         30         40         50         60
MLRLLLRHHF HCLLLCAVWA TPCLASPWST LTANQNPSPP WSKLTYSKPH DAATFYCPFL

        70         80         90        100        110        120
YPSPPRSPLQ FSGFQQVSTG PECRNETLYL LYNREGQTLV ERSSTWVKKV IWYLSGRNQT

       130        140        150        160        170        180
ILQRMPQTAS KPSDGNVQIS VEDAKIFGAH MVPKQTKLLR FVVNDGTRYQ MCVMKLESWA

       190        200        210        220        230        240
HVFRDYSVSF QVRLTFTEAN NQTYTFCTHP NLII*..... .......... ..........
```

Toledo UL132

```
        10         20         30         40         50         60
MPALRGPLRA TFLALVAFGL LLQIDLSDAT NVTSSTKVPT STSNRNNVDN ATSSGPTTGI

        70         80         90        100        110        120
NMTTTHESSV HNVRNNEIMK VLAILFYIVT GTSIFSFIAV LIAVVYSSCC KHPGRFRFAD

       130        140        150        160        170        180
EEAVNLLDDT DDSGGSSPFG SGSRRGSQIP PDFVPRALIS GWKLGTGTRR RRRPRPASA*

       190        200        210        220        230        240
NMILRTSSIS ERMATWTRRS *IPIMGEARL *PSNLTSRTM RRTPSGTTFR CTMN*PPRKW

       250        260        270        280        290        300
KNLRTAPAGR FPN**KLPCN PSRSEIPSTT R......... .......... ..........
```

Toledo UL133

```
        10         20         30         40         50         60
MGCDVHDPSW QCQWGVPTII VAWITCAALG IWCLAGSSAD VSSGPGIAAV VGCSVFMIFL

        70         80         90        100        110        120
CAYLIRYREF FKDSVIDLLT CRWVRYCSCS CKCSCKCISG PCSRCCSACY KETMIYDMVQ

       130        140        150        160        170        180
YGHRRRPGHG DDPDRVICEI VESPPVSAPT VSVPPPSEES HQPVIPPQPP APTSEPKPKK

       190        200        210        220        230        240
GRAKDKPKGR PKDKPPCEPT VSSQPPSQPT AMPGGPPDAP PPAMPQMPPG VAEAVQAAVQ

       250        260        270        280        290        300
AAVAAALQQQ QQHQTGT*.. .......... .......... .......... ..........
```

Fig. 2A

Toledo UL137

```
        10         20         30         40         50         60
MATISTSITP MMGNPTFSGR SSMVTVLCPD LRPSLSLLYS TRAGTAPSTL LRSGRYGVLP

        70         80         90        100        110        120
RATYLHGRLN GGLDRHMHRI HPFWQQCVRR RRTSRG*... .......... ..........
```

Toledo UL138

```
        10         20         30         40         50         60
MDDLPLNVGL PIIGVMLVLI VAILCYLAYH WHDTFKLVRM FLSYRWLIRC CELYGEYERR

        70         80         90        100        110        120
FADLSSLGLG AVRRESDRRY RFSERPDEIL VRWEEVSSQC SYASSRITDR RVGSSSSSSV

       130        140        150        160        170        180
HVASQRNSVP PPDMAVTAPL TDVDLLKPVT GSATQFTTVA MVHYHQEYT* ..........
```

Toledo UL139

```
        10         20         30         40         50         60
MLWILVLFAL AASASETTTG TSSNSSQSTS ATANTTVSTC INASNGSSWT VPQLALLAAS

        70         80         90        100        110        120
GWTLSGLLLL FTCCFCCFWL VRKICSCCGN SSESESKTTH AYTNAAFTSS DATLPMGTTG

       130        140        150        160        170        180
SYTPPQDGSF PPPPR*.... .......... .......... .......... ..........
```

Toledo UL140

```
        10         20         30         40         50         60
MTPAQTNATT TVHPHDAKNG SGGSALPTLV VFGFIVTLLF FLFMLYFWNN DVFRKLLRAL

        70         80         90        100        110        120
GSSAVATAST RGKTRSSTVV HHVVPRATTR VVLTACHRTF FYHPRPMAVL TTRH*.....
```

Fig. 2C

Toledo UL141

```
        10         20         30         40         50         60
MRQVAYRRRR ESSCAVLVHH VGRDGDGEGE AAKKTCKKTG RSVAGIPGEK LRRTVVTTTP

        70         80         90        100        110        120
ARRLSGRHTE QEQAGMRLCE KGKKRIIMCR RESLRTLPWL FWVLLSCPRL LEYSSSSFPF

       130        140        150        160        170        180
ATADIAEKMW AENYETTSPA PVLVAEGEQV TIPCTVMTHS WPMVSIRARF CRSHDGSDEL

       190        200        210        220        230        240
ILDAVKGHRL MNGLQYRLPY ATWNFSQLHL GQIFSLTFNV SMDTAGMYEC VLRNYSHGLI

       250        260        270        280        290        300
MQRFVILTQL ETLSRPDEPC CTPALGRYSL GDQIWSPTPW RLRNHDCGTY RGFQRNYFYI

       310        320        330        340        350        360
GRADAEDCWK PACPDEEPDR CWTVIQRYRL PGDCYRSQPH PPKFLPVTPA PPADIDTGMS

       370        380        390        400        410        420
PWATRGIAAF LGFWSIFTVC FLCYLCYLQC CGRWCPTPGR GRRGGEGYRR LPTYDSYPGV

       430        440        450        460        470        480
RKMKR*.... .......... .......... .......... .......... ..........
```

Toledo UL142

```
        10         20         30         40         50         60
MRIEWVWWLF GYFVSSVGSE RSLSYRYHLE SNSSTNVVCN GNISVFVNGT LGVRYNITVG

        70         80         90        100        110        120
ISSSLLIGHL TIQVLESWFT PWVQNKSYNK QPLGDTETLY NIDSENIHRV SQYFHTRWIK

       130        140        150        160        170        180
SLQENHTCDL TNSTPTYTYQ VNVNNTNYLT LTSSGWQDRL NYTVINSTHF NLTESNITSI

       190        200        210        220        230        240
QKYLNTTCIE RLRNYTLESV YTTTVPQNIT TSQHATTTMH TIPPNTITIQ NTTQSHTVQT

       250        260        270        280        290        300
PSFNDTHNVT KHTLNISYVL SQKTNNTTSP WIYAIPMGAT ATIGAGLYIG KHFTPVKFVY

       310        320        330        340        350        360
EVWRGQ*... .......... .......... .......... .......... ..........
```

Fig. 2D

Toledo UL143

```
        10         20         30         40         50         60
MARSVKTIRI QHIYSPRSSN TLQHMSKKQE SIATITFGRI TCCHPLASIN LMFNGSCTVT

        70         80         90        100        110        120
VKISMGINGS TNVHQLVIVL HLGNRCQPWR QV*....... .......... ..........
```

Toledo UL144

```
        10         20         30         40         50         60
MKPLIMLICF AVILLQLGVT KVCQHNEVQL GNECCPPCGS GQRVTKVCTD YTSVTCTPCP

        70         80         90        100        110        120
NGTYVSGLYN CTDCTQCNVT QVMIRNCTST NNTVCAPKNH TYFSTPGVQH HKQRQQNHTA

       130        140        150        160        170        180
HITVKQGKSG RHTLAWLSLF IFLVGIILLI LYLIAAYRSE RCQQCCSIGK IFYRTL*...
```

Toledo UL145

```
        10         20         30         40         50         60
MCTDPRRTAG WERLTHHASY HANYGAYAVL MATSQRKSLV LHRYSAVTAV ALQLMPVEIV

        70         80         90        100        110        120
RKLDQSDWVR GAWIVSETFP TSDPKGVWSD DDSSMGGSDD *......... ..........
```

Toledo UL146

```
        10         20         30         40         50         60
MRLIFGALII FLAYVYHYEV NGTELRCRCL HRKWPPNKII LGNYWLHRDP RGPGCDKNEH

        70         80         90        100        110        120
LLYPDGRKPP GPGVCLSPDH LFSKWLDKHN DNRWYNVNIT KSPGPRRINI TLIGVRG*..
```

Fig. 2E

Toledo UL147

```
        10         20         30         40         50         60
MVLTWLHHPV SNSHINLLSV RHLSLIAYML LTICPLAVHV LELEDYDRRC RCNNQILLNT

        70         80         90        100        110        120
LPVGTELLKP IAASESCNRQ EVLAILKDKG TKCLNPNAQA VRRHINRLFF RLILDEEQRI

       130        140        150        160        170        180
YDVVSTNIEF GAWPVPTAYK AFLWKYAKRL NYHHFRLRW* .......... ..........
```

Toledo UL148

```
        10         20         30         40         50         60
MLRLLFTLVL LALHGQSVGA SRDYVHVRLL SYRGDPLVFK HTFSGVRRPF TELGWAACRD

        70         80         90        100        110        120
WDSMHCTPFW STDLEQMTDS VRRYSTVSPG KEVTLQLHGN QTVQPSFLSF TCRLQLEPVV

       130        140        150        160        170        180
ENVGLYVAYV VNDGERPQQF FTPQVDVVRF ALYLETLSRI VEPLESGRLA VEFDTPDLAL

       190        200        210        220        230        240
APDLVSSLFV AGHGETDFYM NWTLRRSQTH YLEEMALQVE ILKPRGVRHR AIIHHPKLQP

       250        260        270        280        290        300
GVGLWIDFCV YRYNARLTRG YVRYTLSPKA RLPAKAEGWL VSLDRFIVQY LNTLLITMMA

       310        320        330        340        350        360
AIWARVLITY LVSRRR*... .......... .......... .......... ..........
```

Toledo UL149

```
        10         20         30         40         50         60
MVDQCCYRHL HRSLSGGPDV LYAAAGTQRE QQRLDKSLAA TAPSAVAGPP ADRDVVDHRT

        70         80         90        100        110        120
ETHAYETPRY ATRCLTRYTT PVRSAVRRTT CGKRVASQSP PRSCLVAPQS SPAHPPRHPE

       130        140        150        160        170        180
GG*....... .......... .......... .......... .......... ..........
```

Fig. 2F

Toledo UL150

```
        10         20         30         40         50         60
MQLCSHSISS QRHVASSMHC RSRHQRTPPS ATTHGPCAPT SRILRRLLTT RRFLPRTPSP

        70         80         90        100        110        120
SNTVCCIRRR LHERTIRHSM RCRRRDMASS ASTPVSHTQP LAANHRRSRI TYATTDPTNS

       130        140        150        160        170        180
PTASPAKSDK LEADADPALH RRPASLLRHL FQPCHAQRGT SNRATSQRAS LNAVHHKLCG

       190        200        210        220        230        240
AMISSSCSTT CTPLIMDLPS LSVELSAGHK KKETPTEGGW GGEEGEDDVL ATIRNTLSAP

       250        260        270        280        290        300
TSPAAATTHR LSFPGESTFC LTAVSECSQR RTSTAALTPP PPAVAAAFSF SSTVSETGTF

       310        320        330        340        350        360
PQSTTGRTRV DDTAVVTAGD PRSPVTHVTL LQIFRLRSSL LTSRSGGALR GGEHEAIPKV

       370        380        390        400        410        420
ASLFWTLLKA TQIVEMTHKT PSADSHRNPQ KYTDRPQRLL LTALAIWQRT YNDTRAAHAP

       430        440        450        460        470        480
QVRLLGDILT YRRPQTATAS TKAHTQQQPE EPKGQQIWTQ TAGQAAPHGD EPHSDGELRR

       490        500        510        520        530        540
ESHSAPPTSR TLPDTILAVK RRSVAQRSHV RLDAKPGLNE RDGFRQRLLL PLSGYFRANE

       550        560        570        580        590        600
LRNQQFMGYG TKNGLKNTWL TRPLGVAGGV RETIGERQDR NVADSATQRV FHTLYAALQT

       610        620        630        640        650        660
VRVWYTALGT AWRTSGSRTR ESLFDGPRRR DRQAARLRRL EL*....... ..........
```

Fig. 2G

Toledo UL151

10	20	30	40	50	60
MVFVSGTALG	TGFHRAEGSF	CGCEGRSFFR	TLGTGLGDGG	CAGRRWXRXV	AGTGITLGTG
70	80	90	100	110	120
TRGPGLRDGG	DGGVCGEDGG	LLRRGRGLAG	PAVAGVCGDG	GLLQRRGLRG	QECAXPGGFA
130	140	150	160	170	180
GGHGTGGGGD	STNHTHTQLT	SAVALSEPPL	FFINVLIPPA	YTRNAACSYA	HTLSLHSDML
190	200	210	220	230	240
LRLCTAAADT	SGHRHLPPHM	AHVLRRPASY	VVCSQHGAFF	PARHLHRTPS	AAFAVASTRE
250	260	270	280	290	300
QYATACAVAA	ATWPPRLPHL	FRTPNLWLPT	TDVQGSRTRR	PIPPILQRPR	PPSQTSWKPT
310	320	330	340	350	360
QTQHSIDARP	RCCATSSSPA	TPNAALPTEP	HPRGLP*...	....	....

Fig. 2H

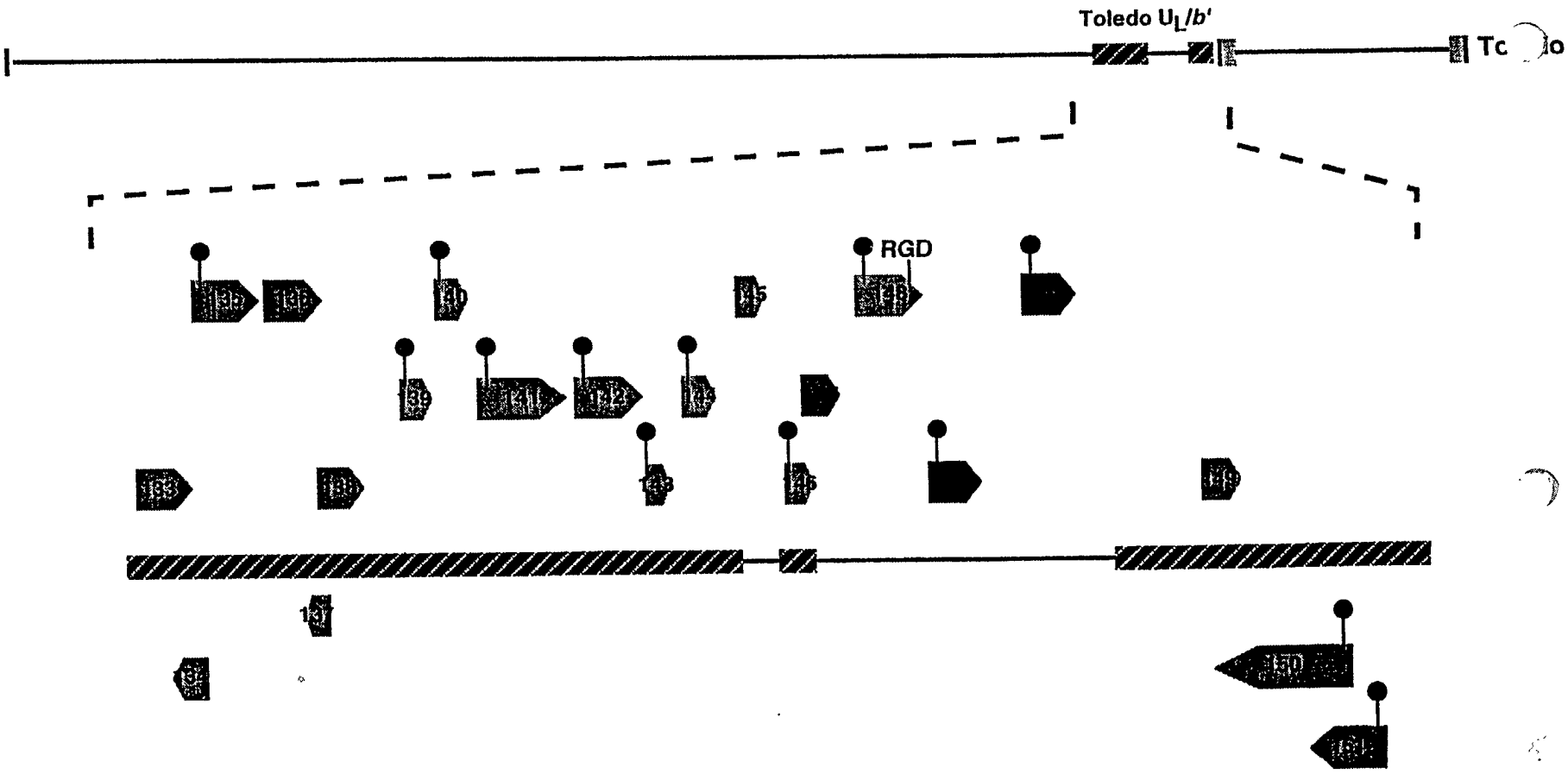


Fig. 3

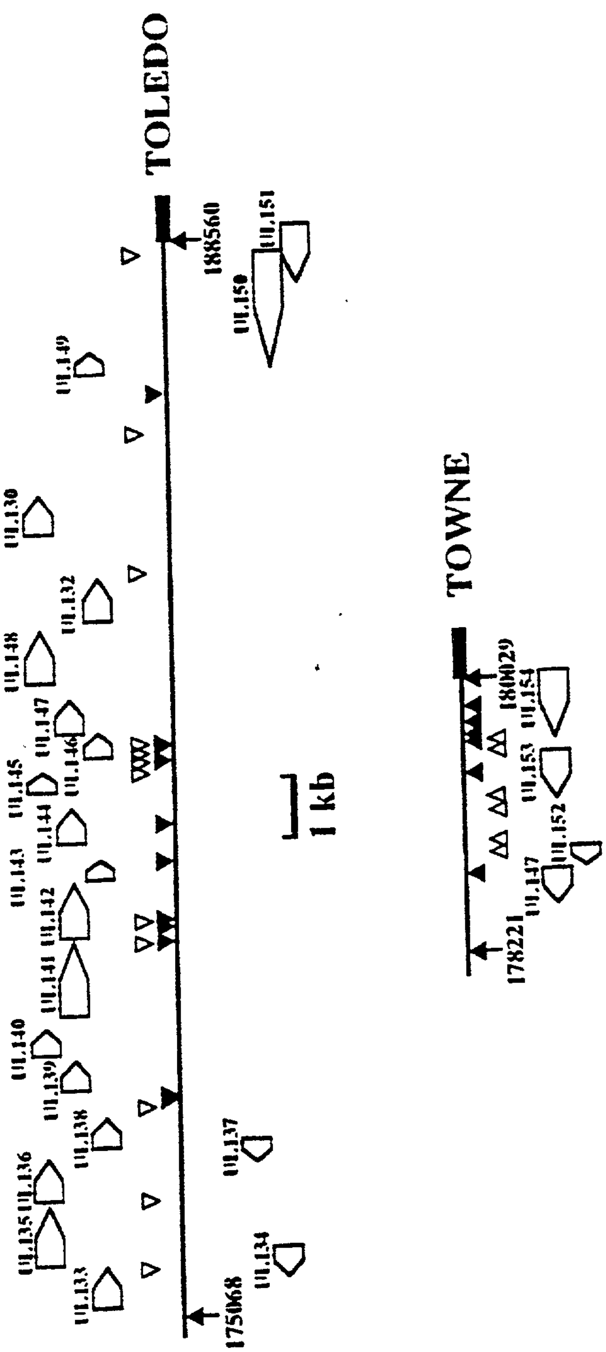
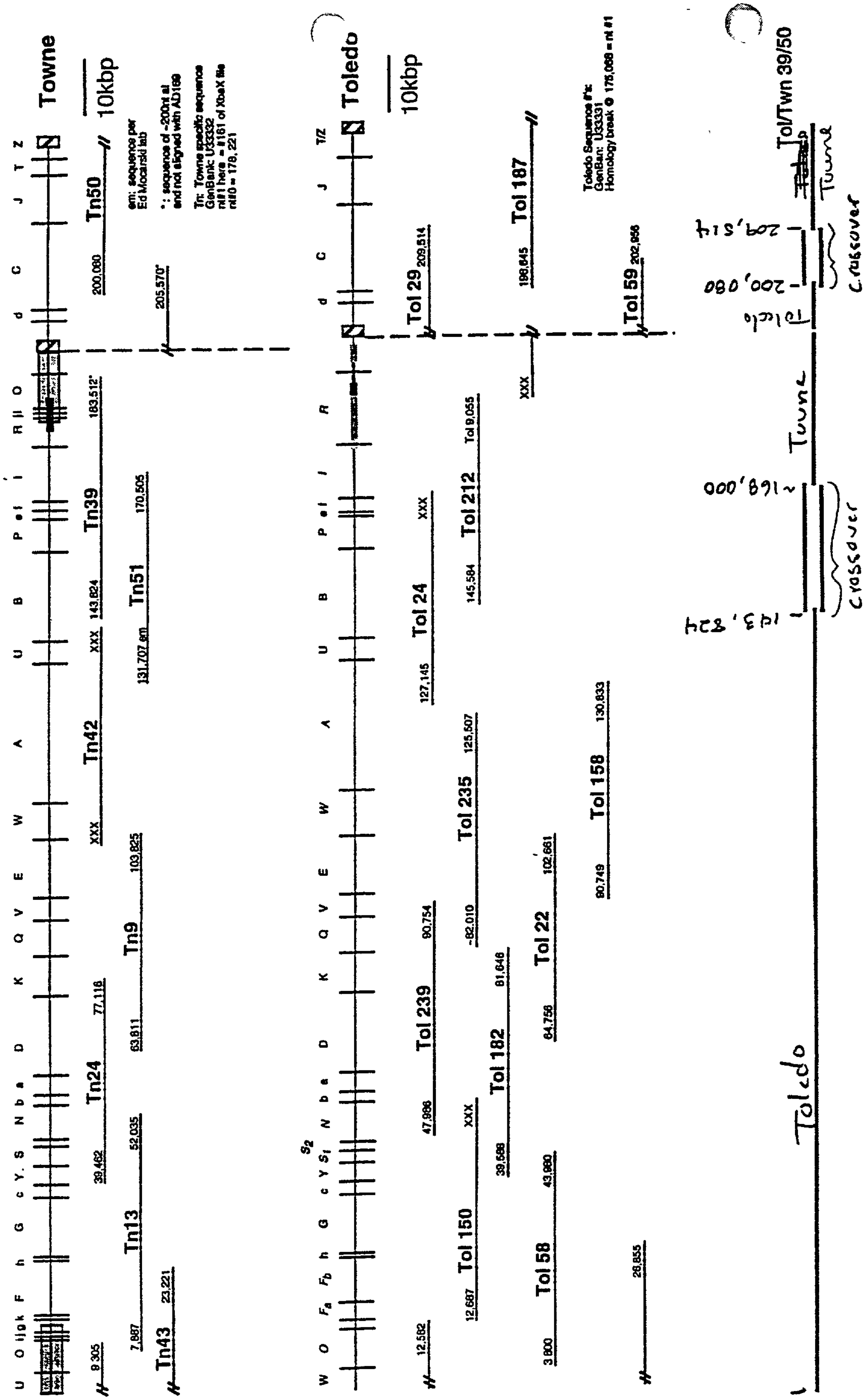


Fig. 4



Towne
10kbp
Toledo
10kbp
Tol/Twn 39/50
Tn50
Tn39
Tn51
Tn42
Tn9
Tn24
Tn13
Tn43
Tol 29
Tol 187
Tol 59
Tol 212
Tol 24
Tol 235
Tol 158
Tol 239
Tol 22
Tol 182
Tol 150
Tol 58
crossover
crossover
Towne
Toledo

Fig. 5

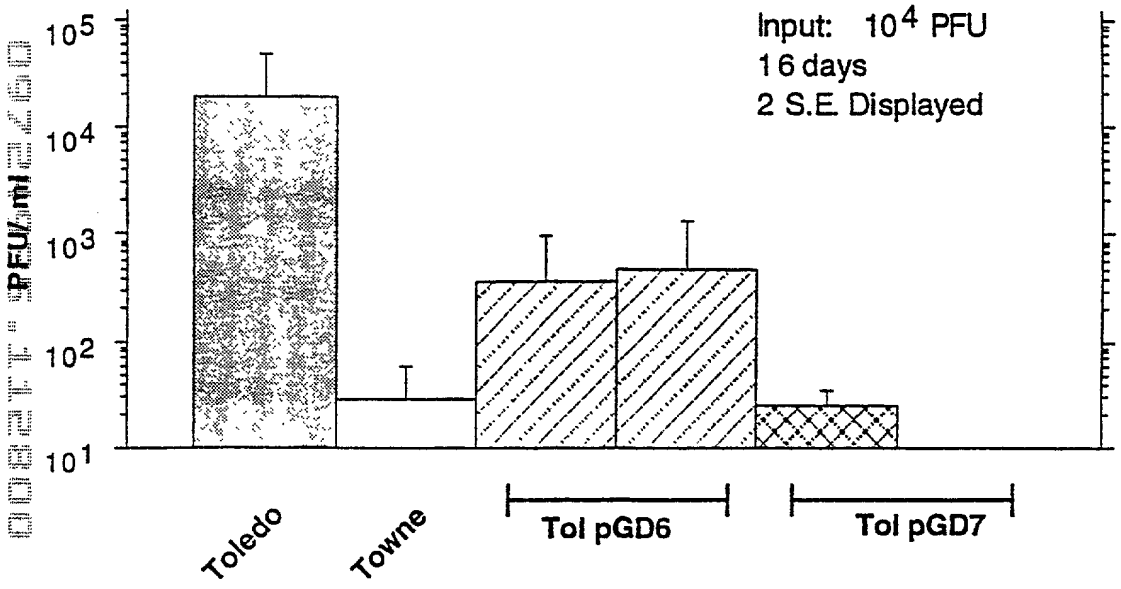

Input: 10^4 PFU
16 days
2 S.E. Displayed
10^5
10^4
10^3
10^2
10^1
Toledo
Towne
Tol pGD6
Tol pGD7

Fig. 6

Clinical Strains of CMV Contain Sequences Homologous to the Toledo U_L/b' Region

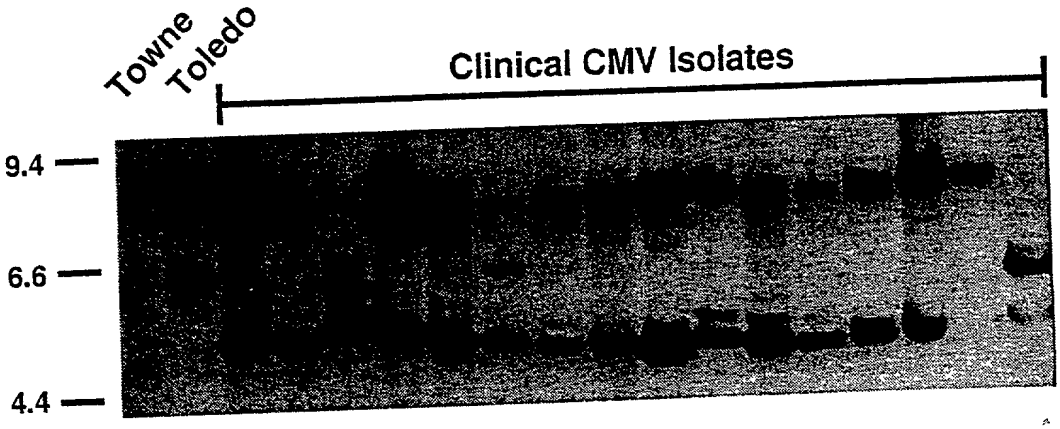


Figure 7

Previous Towne Vaccine Strains Hybridize to the Toledo U_L/*b'* Region

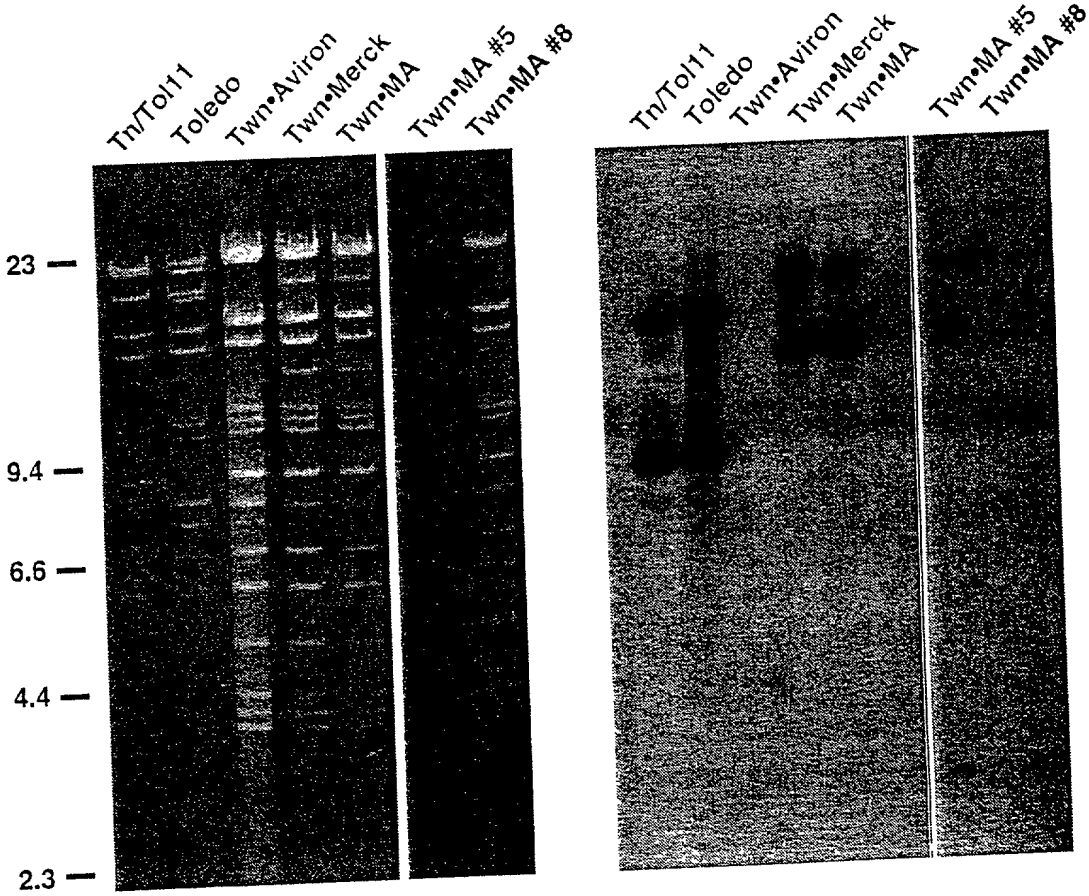


Fig. 5

Cotransfection of Cosmids Regenerates Infectious CMV

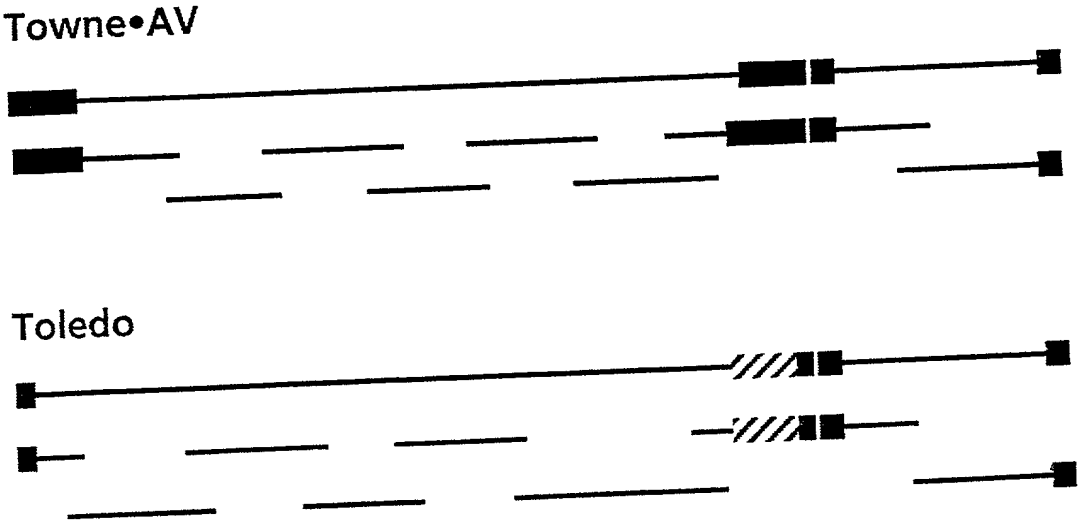


Figure 9

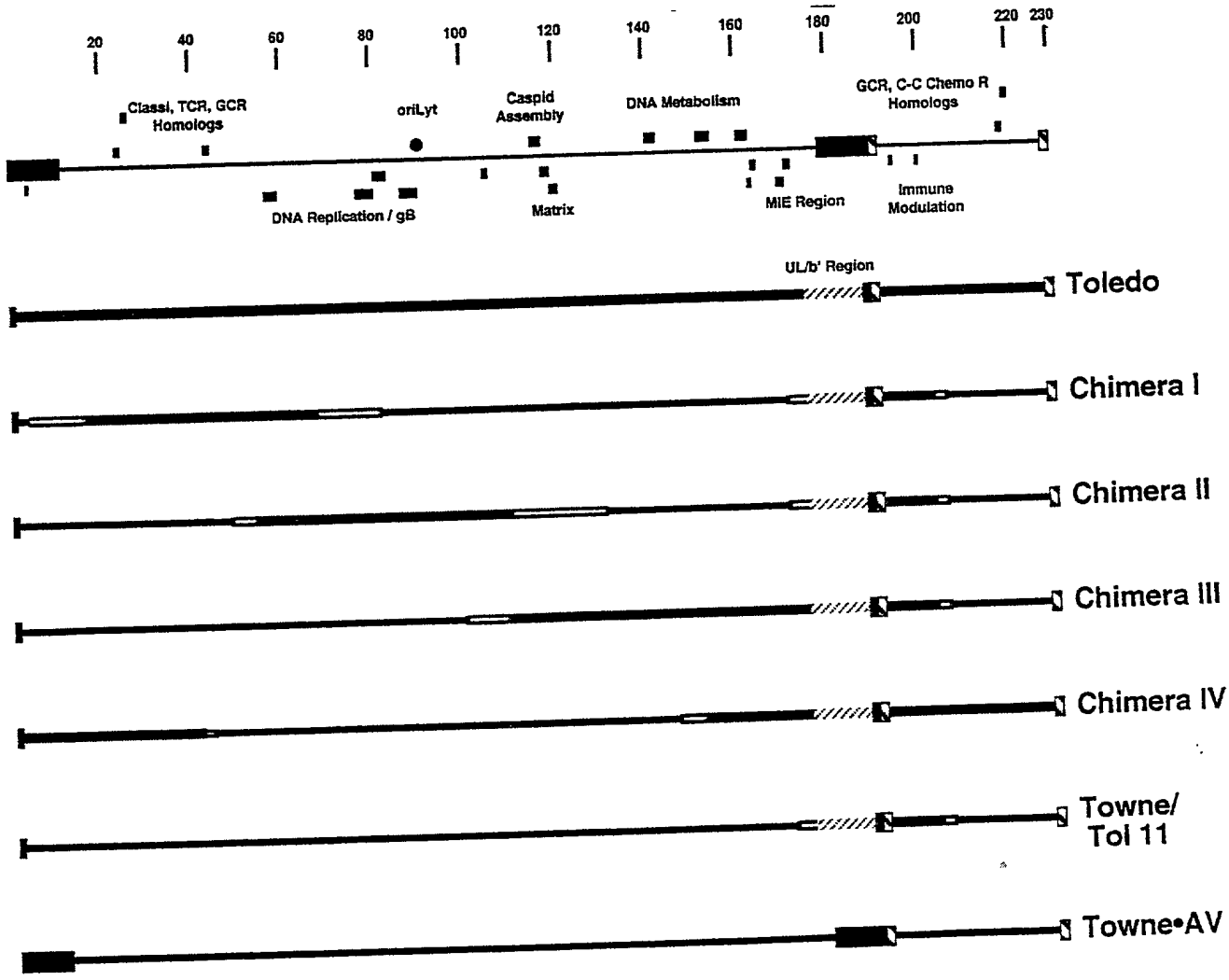

20
40
60
80
100
120
140
160
180
200
220
230
Classl, TCR, GCR Homologs
oriLyt
Caspid Assembly
DNA Metabolism
GCR, C-C Chemo R Homologs
DNA Replication / gB
Matrix
MIE Region
Immune Modulation
UL/b' Region
Toledo
Chimera I
Chimera II
Chimera III
Chimera IV
Towne/ Tol 11
Towne•AV

Figure 10

The Chimeras Replicate in the SCID-hu(thy/liv) Implant

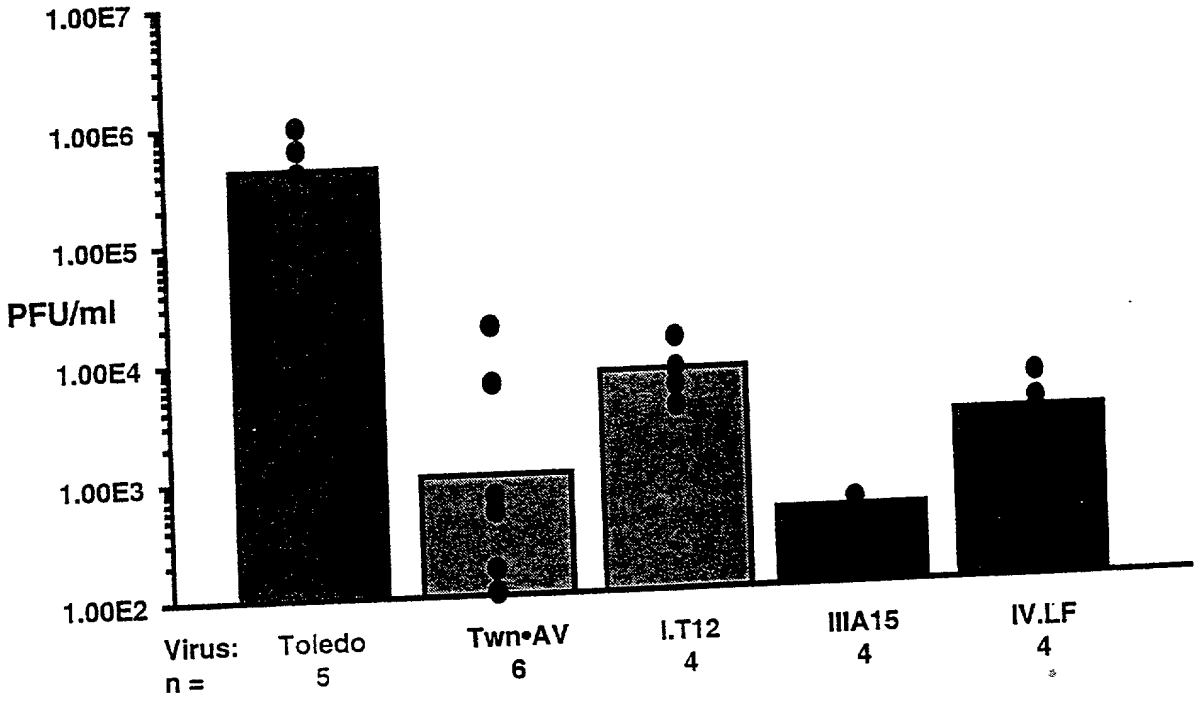


Figure 11

Comparison of the Towne (long) and Towne (short) Genotypes

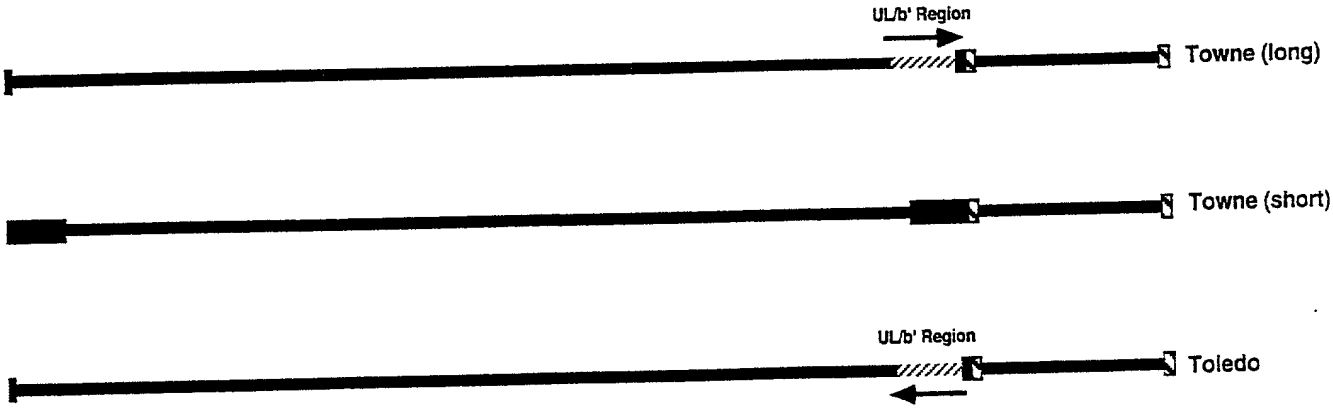


Figure 12

```
        10          20          30          40          50          60
CGCTGTAGGG ATAAATAGTG CGATGGCGTT TGTGGGAGAA CGCAGTAGCG ATGGGTTGCG
GCGACATCCC TATTTATCAC GCTACCGCAA ACACCCTCTT GCGTCATCGC TACCCAACGC

        70          80          90         100         110         120
ACGTGCACGA TCCTTCGTGG CAATGCCAAT GGGGCGTTCC CACGATTATC GTGGCCTGGA
TGCACGTGCT AGGAAGCACC GTTACGGTTA CCCCGCAAGG GTGCTAATAG CACCGGACCT

       130         140         150         160         170         180
TAACATGCGC GGCTTTAGGA ATTTGGTGTT TGGCGGGATC GTCGGCGGAT GTCTCTTCGG
ATTGTACGCG CCGAAATCCT TAAACCACAA ACCGCCCTAG CAGCCGCCTA CAGAGAAGCC

       190         200         210         220         230         240
GACCCGGCAT CGCAGCCGTA GTCGGCTGTT CTGTTTTCAT GATTTTCCTC TGCGCGTATC
CTGGGCCGTA GCGTCGGCAT CAGCCGACAA GACAAAAGTA CTAAAAGGAG ACGCGCATAG

       250         260         270         280         290         300
TCATCCGTTA CCGGGAATTC TTCAAAGACT CCGTAATCGA CCTCCTTACC TGCCGATGGG
AGTAGGCAAT GGCCCTTAAG AAGTTTCTGA GGCATTAGCT GGAGGAATGG ACGGCTACCC

       310         320         330         340         350         360
TTCGCTACTG CAGCTGCAGC TGTAAGTGCA GCTGCAAATG CATCTCGGGC CCCTGTAGCC
AAGCGATGAC GTCGACGTCG ACATTCACGT CGACGTTTAC GTAGAGCCCG GGGACATCGG

       370         380         390         400         410         420
GCTGCTGTTC AGCGTGTTAC AAGGAGACGA TGATTTACGA CATGGTCCAA TACGGTCATC
CGACGACAAG TCGCACAATG TTCCTCTGCT ACTAAATGCT GTACCAGGTT ATGCCAGTAG

       430         440         450         460         470         480
GACGGCGTCC CGGACACGGC GACGATCCCG ACAGGGTGAT CTGCGAGATA GTCGAGAGTC
CTGCCGCAGG GCCTGTGCCG CTGCTAGGGC TGTCCCACTA GACGCTCTAT CAGCTCTCAG

       490         500         510         520         530         540
CCCCGGTTTC GGCGCCGACG GTGTCCGTCC CCCCGCCGTC GGAGGAGTCC CACCAGCCCG
GGGGCCAAAG CCGCGGCTGC CACAGGCAGG GGGGCGGCAG CCTCCTCAGG GTGGTCGGGC

       550         560         570         580         590         600
TCATCCCACC GCAGCCGCCA GCACCGACAT CGGAACCCAA ACCGAAGAAA GGTAGGGCGA
AGTAGGGTGG CGTCGGCGGT CGTGGCTGTA GCCTTGGGTT TGGCTTCTTT CCATCCCGCT

       610         620         630         640         650         660
AAGATAAACC GAAGGGTAGA CCGAAAGACA AACCTCCGTG CGAACCGACG GTGAGTTCAC
TTCTATTTGG CTTCCCATCT GGCTTTCTGT TTGGAGGCAC GCTTGGCTGC CACTCAAGTG

       670         680         690         700         710         720
AACCACCGTC GCAGCCGACG GCAATGCCCG GCGGTCCGCC CGACGCGCCT CCCCCCGCCA
TTGGTGGCAG CGTCGGCTGC CGTTACGGGC CGCCAGGCGG GCTGCGCGGA GGGGGGCGGT

       730         740         750         760         770         780
TGCCGCAGAT GCCACCCGGC GTGGCCGAGG CGGTACAAGC TGCCGTGCAG GCGGCCGTGG
ACGGCGTCTA CGGTGGGCCG CACCGGCTCC GCCATGTTCG ACGGCACGTC CGCCGGCACC

       790         800         810         820         830         840
CCGCGGCTCT ACAACAACAG CAGCAGCATC AGACCGGAAC GTAACCCGCC CCCGGTGCGA
GGCGCCGAGA TGTTGTTGTC GTCGTCGTAG TCTGGCCTTG CATTGGGCGG GGGCCACGCT

       850         860         870         880         890         900
TAAGGAATTT TCCGACTTGG CGCACATCTC CTTCCTCAAT GTTTGGACAA TAAACACATT
ATTCCTTAAA AGGCTGAACC GCGTGTAGAG GAAGGAGTTA CAAACCTGTT ATTTGTGTAA

       910         920         930         940         950         960
CCTTGCCAAA AAATGACGTT TCCAGAAATC CAAGGCATAA ATGTCCGTAC ACCGGCCCTT
GGAACGGTTT TTTACTGCAA AGGTCTTTAG GTTCCGTATT TACAGGCATG TGGCCGGGAA

       970         980         990        1000        1010        1020
CCCAACACGG AGTTTGAGAT TCCAAGCAGG AGAGAAGATC ATGGTGTGGA TATGGCTCGG
```

SEQ NO 1
18318 bases

Fig. 1A

```
      1030        1040        1050        1060        1070        1080
CATCGGGCTC CTCGGCGGTA CCGGACTGGC TTCCCTGGTC CTGGCCATTT CCTTATTTAC
GTAGCCCGAG GAGCCGCCAT GGCCTGACCG AAGGGACCAG GACCGGTAAA GGAATAAATG

      1090        1100        1110        1120        1130        1140
CCAGCGCCGA GGCCGCAAGC GATCCGACGA GACTTCGTCG CGAGGCCGGC TCCCGGGTGC
GGTCGCGGCT CCGGCGTTCG CTAGGCTGCT CTGAAGCAGC GCTCCGGCCG AGGGCCCACG

      1150        1160        1170        1180        1190        1200
TGCTTCTGAT AAGCGTGGTG CCTGCGCGTG CTGCTATCGA AATCCGAAAG AAGACGTCGT
ACGAAGACTA TTCGCACCAC GGACGCGCAC GACGATAGCT TTAGGCTTTC TTCTGCAGCA

      1210        1220        1230        1240        1250        1260
CGAGCCGCTG GATCTGGAAC TGGGGCTCAT GCGGGTGGAC ACCCACCCGC CGACGCCGCA
GCTCGGCGAC CTAGACCTTG ACCCCGAGTA CGCCCACCTG TGGGTGGGCG GCTGCGGCGT

      1270        1280        1290        1300        1310        1320
GGTGCCGCGG TGTACGTCGC TCTACATAGG AGAGGATGGT CTGCCGATAG ATAAACCCGA
CCACGGCGCC ACATGCAGCG AGATGTATCC TCTCCTACCA GACGGCTATC TATTTGGGCT

      1330        1340        1350        1360        1370        1380
GTTTCCTCCG GCGCGGTTCG AGATCCCCGA CGTATCCACG CCGGGAACGC CGACCAGCAT
CAAAGGAGGC CGCGCCAAGC TCTAGGGGCT GCATAGGTGC GGCCCTTGCG GCTGGTCGTA

      1390        1400        1410        1420        1430        1440
CGGCCGATCT CCGTCGCATT GCTCCTCGTC GAGCTCTTTG TCGTCCTCGA CCAGCGTCGA
GCCGGCTAGA GGCAGCGTAA CGAGGAGCAG CTCGAGAAAC AGCAGGAGCT GGTCGCAGCT

      1450        1460        1470        1480        1490        1500
CACGGTGCTG TATCAGCCGC CGCCATCCTG GAAGCCACCT CCGCCGCCCG GGCGCAAGAA
GTGCCACGAC ATAGTCGGCG GCGGTAGGAC CTTCGGTGGA GGCGGCGGGC CCGCGTTCTT

      1510        1520        1530        1540        1550        1560
GCGGCCGCCT ACGCCGCCGG TCCGGGCCCC CACCACGCGG CTGTCGTCGC ACAGACCCCC
CGCCGGCGGA TGCGGCGGCC AGGCCCGGGG GTGGTGCGCC GACAGCAGCG TGTCTGGGGG

      1570        1580        1590        1600        1610        1620
GACGCCGATA CCCGCGCCGC GTAAGAACCT GAGCACGCCG CCCACCAAGA AAACGCCGCC
CTGCGGCTAT GGGCGCGGCG CATTCTTGGA CTCGTGCGGC GGGTGGTTCT TTTGCGGCGG

      1630        1640        1650        1660        1670        1680
GCCCACGAAA CCCAAGCCGG TCGGCTGGAC ACCGCCGGTG ACACCCAGGC CCTTCCCGAA
CGGGTGCTTT GGGTTCGGCC AGCCGACCTG TGGCGGCCAC TGTGGGTCCG GGAAGGGCTT

      1690        1700        1710        1720        1730        1740
AACGCCGACG CCACAAAAGC CGCCGCGGAA TCCGAGACTA CCGCGCACCG TCGGTCTGGA
TTGCGGCTGC GGTGTTTTCG GCGGCGCCTT AGGCTCTGAT GGCGCGTGGC AGCCAGACCT

      1750        1760        1770        1780        1790        1800
GAATCTCTCG AAGGTGGGAC TCTCGTGTCC CTGTCCCCGA CCCCGCACGC CGACGGAGCC
CTTAGAGAGC TTCCACCCTG AGAGCACAGG GACAGGGGCT GGGGCGTGCG GCTGCCTCGG

      1810        1820        1830        1840        1850        1860
GACCACGCTG CCTATCGTGT CGGTTTCCGA GCTAGCCCCG CCTCCTCGAT GGTCGGACAT
CTGGTGCGAC GGATAGCACA GCCAAAGGCT CGATCGGGGC GGAGGAGCTA CCAGCCTGTA

      1870        1880        1890        1900        1910        1920
CGAGGAACTC TTGGAACAGG CGGTGCAGAG CGTCATGAAG GACGCCGAGT CGATGCAGAT
GCTCCTTGAG AACCTTGTCC GCCACGTCTC GCAGTACTTC CTGCGGCTCA GCTACGTCTA

      1930        1940        1950        1960        1970        1980
GACCTGAGAC CGAAAGAGCG AGCGCGTCCG TTGTACAGTT GTATAGCAGC ACACGCCTTC
CTGGACTCTG GCTTTCTCGC TCGCGCAGGC AACATGTCAA CATATCGTCG TGTGCGGAAG

      1990        2000        2010        2020        2030        2040
CCTCTTTTTC ACCGCAGCTA AGAGAGAGAA AGAGAGTATG TCAGTCAAGG GCGTGGAGAT
GGAGAAAAAG TGGCGTCGAT TCTCTCTCTT TCTCTCATAC AGTCAGTTCC CGCACCTCTA
```

Fig. 1B

```
      2050        2060        2070        2080        2090        2100
GCCAGAAATG ACGTGGGACT TGGACGTTAG AAATAAATGG CGGCGTCGAA AGGCCCTGAG
CGGTCTTTAC TGCACCCTGA ACCTGCAATC TTTATTTACC GCCGCAGCTT TCCGGGACTC

      2110        2120        2130        2140        2150        2160
TCGCATTCAC CGGTTCTGGG AATGTCGGCT ACGGGTGTGG TGGCTGAGTG ACGCCGGCGT
AGCGTAAGTG GCCAAGACCC TTACAGCCGA TGCCCACACC ACCGACTCAC TGCGGCCGCA

      2170        2180        2190        2200        2210        2220
AAGAGAAACC GACCCACCGC GTCCCCGACG CCGCCCGACT TGGATGACCG CGGTGTTTCA
TTCTCTTTGG CTGGGTGGCG CAGGGGCTGC GGCGGGCTGA ACCTACTGGC GCCACAAAGT

      2230        2240        2250        2260        2270        2280
CGTTATCTGT GCCGTTTTGC TTACGCTTAT GATTATGGCC ATCGGCGCGC TCATCGCGTA
GCAATAGACA CGGCAAAACG AATGCGAATA CTAATACCGG TAGCCGCGCG AGTAGCGCAT

      2290        2300        2310        2320        2330        2340
CTTAAGATAT TACCACCAGG ACAGTTGGCG AGACATGCTC CACGATCTAT TTTGCGGCTG
GAATTCTATA ATGGTGGTCC TGTCAACCGC TCTGTACGAG GTGCTAGATA AAACGCCGAC

      2350        2360        2370        2380        2390        2400
TCATTATCCC GAGAAGTGCC GTCGGCACCA CGAGCGGCAG AGAAGGAGAC GGCAAGCCAT
AGTAATAGGG CTCTTCACGG CAGCCGTGGT GCTCGCCGTC TCTTCCTCTG CCGTTCGGTA

      2410        2420        2430        2440        2450        2460
GGATGTGCCC GACCCGGAAC TCGGCGACCC GGCCCGCCGG CCGTTGAACG GAGCTATGTA
CCTACACGGG CTGGGCCTTG AGCCGCTGGG CCGGGCGGCC GGCAACTTGC CTCGATACAT

      2470        2480        2490        2500        2510        2520
CTACGGCAGC GGCTGTCGCT TCGACACGGT GGAAATGGTG GACGAGACGA GACCCGCGCC
GATGCCGTCG CCGACAGCGA AGCTGTGCCA CCTTTACCAC CTGCTCTGCT CTGGGCGCGG

      2530        2540        2550        2560        2570        2580
GCCGGCGCTG TCATCGCCCG AAACCGGCGA CGATAGCAAC GACGACGCGG TTGCCGGCGG
CGGCCGCGAC AGTAGCGGGC TTTGGCCGCT GCTATCGTTG CTGCTGCGCC AACGGCCGCC

      2590        2600        2610        2620        2630        2640
AGGTGCTGGC GGGGTAACAT CACCCGCGAC TCGTACGACG TCGCCGAACG CACTGCTGCC
TCCACGACCG CCCCATTGTA GTGGGCGCTG AGCATGCTGC AGCGGCTTGC GTGACGACGG

      2650        2660        2670        2680        2690        2700
AGAATGGATG GATGCGGTGC ATGTGGCGGT CCAAGCCGCC GTTCAAGCGA CCGTGCAAGT
TCTTACCTAC CTACGCCACG TACACCGCCA GGTTCGGCGG CAAGTTCGCT GGCACGTTCA

      2710        2720        2730        2740        2750        2760
AAGTGGCCCG CGGGAGAACG CCGTATCTCC CGCTACGTAA GAGGGTTGAG GGGGCCGTTC
TTCACCGGGC GCCCTCTTGC GGCATAGAGG GCGATGCATT CTCCCAACTC CCCCGGCAAG

      2770        2780        2790        2800        2810        2820
CCGCGCGAGT GCTGTACAAA AGAGAGAGAC TGGGACGTAG ATCCGGACAG AGGACGGTCA
GGCGCGCTCA CGACATGTTT TCTCTCTCTG ACCCTGCATC TAGGCCTGTC TCCTGCCAGT

      2830        2840        2850        2860        2870        2880
CCATGGACGA TCTGCCGCTG AATGTCGGGT TACCCATCAT CGGCGTGATG CTCGTGCTGA
GGTACCTGCT AGACGGCGAC TTACAGCCCA ATGGGTAGTA GCCGCACTAC GAGCACGACT

      2890        2900        2910        2920        2930        2940
TCGTGGCCAT CCTCTGCTAT CTGGCTTACC ACTGGCACGA CACCTTCAAA CTGGTGCGCA
AGCACCGGTA GGAGACGATA GACCGAATGG TGACCGTGCT GTGGAAGTTT GACCACGCGT

      2950        2960        2970        2980        2990        3000
TGTTTCTGAG CTACCGCTGG CTGATCCGCT GTTGCGAGCT GTACGGGGAG TACGAGCGCC
ACAAAGACTC GATGGCGACC GACTAGGCGA CAACGCTCGA CATGCCCCTC ATGCTCGCGG

      3010        3020        3030        3040        3050        3060
GGTTCGCGGA CCTGTCGTCT CTGGGCCTCG GCGCCGTACG GCGGGAGTCG GACAGACGAT
CCAAGCGCCT GGACAGCAGA GACCCGGAGC CGCGGCATGC CGCCCTCAGC CTGTCTGCTA
```

Fig. 1C

```
      3070        3080        3090        3100        3110        3120
ACCGTTTCTC CGAACGGCCC GACGAGATCT TGGTCCGTTG GGAGGAAGTG TCTTCCCAGT
TGGCAAAGAG GCTTGCCGGG CTGCTCTAGA ACCAGGCAAC CCTCCTTCAC AGAAGGGTCA

      3130        3140        3150        3160        3170        3180
GCAGCTACGC GTCGTCGCGG ATAACAGACC GCCGTGTGGG TTCATCGTCT TCGTCGTCGG
CGTCGATGCG CAGCAGCGCC TATTGTCTGG CGGCACACCC AAGTAGCAGA AGCAGCAGCC

      3190        3200        3210        3220        3230        3240
TCCACGTCGC TAGCCAGAGA AACAGCGTGC CTCCGCCGGA CATGGCGGTG ACGGCGCCGC
AGGTGCAGCG ATCGGTCTCT TTGTCGCACG GAGGCGGCCT GTACCGCCAC TGCCGCGGCG

      3250        3260        3270        3280        3290        3300
TGACCGACGT CGATCTGTTG AAACCCGTGA CGGGATCCGC GACGCAGTTC ACCACCGTAG
ACTGGCTGCA GCTAGACAAC TTTGGGCACT GCCCTAGGCG CTGCGTCAAG TGGTGGCATC

      3310        3320        3330        3340        3350        3360
CCATGGTACA TTATCATCAA GAGTACACGT GAATGAGAAA AAGAAAAAAG AGGGGAGCGG
GGTACCATGT AATAGTAGTT CTCATGTGCA CTTACTCTTT TTCTTTTTTC TCCCCTCGCC

      3370        3380        3390        3400        3410        3420
ATCGCGATAA TGTCGCTTTG ACATTCTCTG CTCGATCTAC TCAGCGTCTG CACGAAACGG
TAGCGCTATT ACAGCGAAAC TGTAAGAGAC GAGCTAGATG AGTCGCAGAC GTGCTTTGCC

      3430        3440        3450        3460        3470        3480
CATCCGCACG GAGGCGAGCC CAAGCGTATC TGCAGCAAGC GGTTCTTTCC CTCGGTGATG
GTAGGCGTGC CTCCGCTCGG GTTCGCATAG ACGTCGTTCG CCAAGAAAGG GAGCCACTAC

      3490        3500        3510        3520        3530        3540
GTGGCAGCAT CGGTGGCGGG AGCTTGTTCG GACGATGGAC GGTGAGGAGT CCCTGGCGAT
CACCGTCGTA GCCACCGCCC TCGAACAAGC CTGCTACCTG CCACTCCTCA GGGACCGCTA

      3550        3560        3570        3580        3590        3600
CAGGCGGCTC CCGGGTGTGG AGTTCAACGG GTGGTAATGG TGGCGGTGAT CGGTGTTAGA
GTCCGCCGAG GGCCCACACC TCAAGTTGCC CACCATTACC ACCGCCACTA GCCACAATCT

      3610        3620        3630        3640        3650        3660
AAACGGTGGC CCTGGCAAAC ATATATCTAC TGTAAACCCT CTGCTCTGTT AATAAAAAGC
TTTGCCACCG GGACCGTTTG TATATAGATG ACATTTGGGA GACGAGACAA TTATTTTTCG

      3670        3680        3690        3700        3710        3720
ACACTTTTCA CATGAGTTCG TAATTTTATT GTGTAGTGGA AATTTTTACG TCATTGGGAA
TGTGAAAAGT GTACTCAAGC ATTAAAATAA CACATCACCT TTAAAAATGC AGTAACCCTT

      3730        3740        3750        3760        3770        3780
ACCCCAGAAT GAAAGAGTAT AATGTGCATA TCACCGGGGG TTCCCTGTCA GTACGAATGT
TGGGGTCTTA CTTTCTCATA TTACACGTAT AGTGGCCCCC AAGGGACAGT CATGCTTACA

      3790        3800        3810        3820        3830        3840
ACACAACGCG GGTTACATTA CGATAAACTT TCCGGTAAAA CGATGCCGAT ACAGCGTGTA
TGTGTTGCGC CCAATGTAAT GCTATTTGAA AGGCCATTTT GCTACGGCTA TGTCGCACAT

      3850        3860        3870        3880        3890        3900
TAACGCTGAT TGTTACGACA AACGAGTTGG TATATCCATT ATATAGTAAC GAACATGCTG
ATTGCGACTA ACAATGCTGT TTGCTCAACC ATATAGGTAA TATATCATTG CTTGTACGAC

      3910        3920        3930        3940        3950        3960
TGGATATTAG TTTTATTTGC ACTCGCCGCA TCGGCGAGTG AAACCACTAC AGGTACCAGC
ACCTATAATC AAAATAAACG TGAGCGGCGT AGCCGCTCAC TTTGGTGATG TCCATGGTCG

      3970        3980        3990        4000        4010        4020
TCTAATTCCA GTCAATCTAC TAGTGCTACC GCCAACACGA CCGTATCGAC ATGTATTAAT
AGATTAAGGT CAGTTAGATG ATCACGATGG CGGTTGTGCT GGCATAGCTG TACATAATTA

      4030        4040        4050        4060        4070        4080
GCCTCTAACG GCAGTAGCTG GACAGTACCA CAGCTCGCGC TGCTTGCCGC TAGCGGCTGG
CGGAGATTGC CGTCATCGAC CTGTCATGGT GTCGAGCGCG ACGAACGGCG ATCGCCGACC
```

Fig. 1D

```
      4090        4100        4110        4120        4130        4140
ACATTATCTG GACTCCTTCT CTTATTTACC TGCTGCTTTT GCTGCTTTTG GCTAGTACGT
TGTAATAGAC CTGAGGAAGA GAATAAATGG ACGACGAAAA CGACGAAAAC CGATCATGCA

      4150        4160        4170        4180        4190        4200
AAAATCTGCA GCTGCTGCGG CAACTCCTCC GAGTCAGAGA GCAAAACAAC CCACGCGTAC
TTTTAGACGT CGACGACGCC GTTGAGGAGG CTCAGTCTCT CGTTTTGTTG GGTGCGCATG

      4210        4220        4230        4240        4250        4260
ACCAATGCCG CATTCACTTC TTCCGACGCA ACGTTACCCA TGGGCACTAC AGGGTCGTAC
TGGTTACGGC GTAAGTGAAG AAGGCTGCGT TGCAATGGGT ACCCGTGATG TCCCAGCATG

      4270        4280        4290        4300        4310        4320
ACTCCCCCAC AGGACGGCTC ATTTCCACCT CCGCCTCGGT GACGTAGGCT AAACCGAAAC
TGAGGGGGTG TCCTGCCGAG TAAAGGTGGA GGCGGAGCCA CTGCATCCGA TTTGGCTTTG

      4330        4340        4350        4360        4370        4380
CCACGTTGAA CCTAACGCGG TTTCGGAAGG CCTGAGACGT CACTTTCACA ATGACGTCCG
GGTGCAACTT GGATTGCGCC AAAGCCTTCC GGACTCTGCA GTGAAAGTGT TACTGCAGGC

      4390        4400        4410        4420        4430        4440
TATACACGTT CATCATAAAA CACCGTAGAG GCTAAGGCTT CGGTAGGGAG AGACCTCAAC
ATATGTGCAA GTAGTATTTT GTGGCATCTC CGATTCCGAA GCCATCCCTC TCTGGAGTTG

      4450        4460        4470        4480        4490        4500
TGTTCCTGAT GAGCACCCGT GCTCTCATCT CTTCAGACTT GTCATGACCC CCGCTCAGAC
ACAAGGACTA CTCGTGGGCA CGAGAGTAGA GAAGTCTGAA CAGTACTGGG GGCGAGTCTG

      4510        4520        4530        4540        4550        4560
TAACGCGACT ACCACCGTGC ACCCGCACGA CGCAAAAAAC GGCAGCGGCG GTAGTGCCCT
ATTGCGCTGA TGGTGGCACG TGGGCGTGCT GCGTTTTTTG CCGTCGCCGC CATCACGGGA

      4570        4580        4590        4600        4610        4620
GCCGACCCTC GTCGTTTTCG GCTTTATCGT TACGCTACTT TTCTTTCTCT TTATGCTCTA
CGGCTGGGAG CAGCAAAAGC CGAAATAGCA ATGCGATGAA AAGAAAGAGA AATACGAGAT

      4630        4640        4650        4660        4670        4680
CTTTTGGAAC AACGACGTGT TCCGTAAGCT GCTCCGTGCG CTTGGATCCA GCGCTGTTGC
GAAAACCTTG TTGCTGCACA AGGCATTCGA CGAGGCACGC GAACCTAGGT CGCGACAACG

      4690        4700        4710        4720        4730        4740
GACCGCTTCG ACGCGTGGCA AGACGAGGTC ATCTACCGTC GTCCATCACG TCGTTCCCAG
CTGGCGAAGC TGCGCACCGT TCTGCTCCAG TAGATGGCAG CAGGTAGTGC AGCAAGGGTC

      4750        4760        4770        4780        4790        4800
AGCGACGACG AGAGTCGTAC TAACAGCGTG TCATCGTACG TTCTTTTATC ACCCGCGTCC
TCGCTGCTGC TCTCAGCATG ATTGTCGCAC AGTAGCATGC AAGAAAATAG TGGGCGCAGG

      4810        4820        4830        4840        4850        4860
GATGGCGGTT TTGACAACCC GGCACTGACA GAGGCCGTCG ACAGCGTGGA CGACTGGGCG
CTACCGCCAA AACTGTTGGG CCGTGACTGT CTCCGGCAGC TGTCGCACCT GCTGACCCGC

      4870        4880        4890        4900        4910        4920
ACCACCTCGG TTTTCTACGC CACGTCCGAC GAAACGGCGG ACGCCGAGCG CCGAGACTCG
TGGTGGAGCC AAAAGATGCG GTGCAGGCTG CTTTGCCGCC TGCGGCTCGC GGCTCTGAGC

      4930        4940        4950        4960        4970        4980
CAGCAACTGC TCATCGAGCT TCCGCCGGAG CCGCTCCCGC CCGACGTGGT GGCGGCCATG
GTCGTTGACG AGTAGCTCGA AGGCGGCCTC GGCGAGGGCG GGCTGCACCA CCGCCGGTAC

      4990        5000        5010        5020        5030        5040
CAGAAAGCAG TGAAACGCGC TGTACAGAAC GCACTACGAC ACAGCCACGA CTCTTGGCAG
GTCTTTCGTC ACTTTGCGCG ACATGTCTTG CGTGATGCTG TGTCGGTGCT GAGAACCGTC

      5050        5060        5070        5080        5090        5100
CTTCATCAGA CCCTGTGACG CCAGATGAAC GTTCCTTCTT AAACATCCGA GGTAGCAATG
GAAGTAGTCT GGGACACTGC GGTCTACTTG CAAGGAAGAA TTTGTAGGCT CCATCGTTAC
```

Fig. 1E

```
      5110        5120        5130        5140        5150        5160
AGACAGGTCG CGTACCGCCG GCGACGCGAG AGTTCCTGCG CGGTGCTGGT CCACCACGTC
TCTGTCCAGC GCATGGCGGC CGCTGCGCTC TCAAGGACGC GCCACGACCA GGTGGTGCAG

      5170        5180        5190        5200        5210        5220
GGCCGCGACG GCGACGGCGA GGGGGAGGCA GCAAAAAAGA CCTGCAAAAA AACCGGACGC
CCGGCGCTGC CGCTGCCGCT CCCCCTCCGT CGTTTTTTCT GGACGTTTTT TTGGCCTGCG

      5230        5240        5250        5260        5270        5280
TCAGTTGCGG GCATCCCGGG CGAGAAGCTG CGTCGCACGG TGGTCACCAC CACGCCGGCC
AGTCAACGCC CGTAGGGCCC GCTCTTCGAC GCAGCGTGCC ACCAGTGGTG GTGCGGCCGG

      5290        5300        5310        5320        5330        5340
CGACGTTTGA GCGGCCGACA CACGGAGCAG GAGCAGGCGG GCATGCGTCT CTGTGAAAAA
GCTGCAAACT CGCCGGCTGT GTGCCTCGTC CTCGTCCGCC CGTACGCAGA GACACTTTTT

      5350        5360        5370        5380        5390        5400
GGGAAGAAAA GAATCATCAT GTGCCGCCGG GAGTCGCTCC GAACTCTGCC GTGGCTGTTC
CCCTTCTTTT CTTAGTAGTA CACGGCGGCC CTCAGCGAGG CTTGAGACGG CACCGACAAG

      5410        5420        5430        5440        5450        5460
TGGGTGCTGT TGAGCTGCCC GCGACTCCTC GAATATTCTT CCTCTTCGTT CCCCTTCGCC
ACCCACGACA ACTCGACGGG CGCTGAGGAG CTTATAAGAA GGAGAAGCAA GGGGAAGCGG

      5470        5480        5490        5500        5510        5520
ACCGCTGACA TTGCCGAAAA GATGTGGGCC GAGAATTATG AGACCACGTC GCCGGCGCCG
TGGCGACTGT AACGGCTTTT CTACACCCGG CTCTTAATAC TCTGGTGCAG CGGCCGCGGC

      5530        5540        5550        5560        5570        5580
GTGTTGGTCG CCGAGGGAGA GCAAGTTACC ATCCCCTGCA CGGTCATGAC ACACTCCTGG
CACAACCAGC GGCTCCCTCT CGTTCAATGG TAGGGGACGT GCCAGTACTG TGTGAGGACC

      5590        5600        5610        5620        5630        5640
CCCATGGTCT CCATTCGCGC ACGTTTCTGT CGTTCCCACG ACGGCAGCGA CGAGCTCATC
GGGTACCAGA GGTAAGCGCG TGCAAAGACA GCAAGGGTGC TGCCGTCGCT GCTCGAGTAG

      5650        5660        5670        5680        5690        5700
CTGGACGCCG TCAAAGGCCA TCGGCTGATG AACGGACTCC AGTACCGCCT GCCGTACGCC
GACCTGCGGC AGTTTCCGGT AGCCGACTAC TTGCCTGAGG TCATGGCGGA CGGCATGCGG

      5710        5720        5730        5740        5750        5760
ACTTGGAATT TCTCGCAATT GCATCTCGGC CAAATATTCT CGCTTACTTT TAACGTATCG
TGAACCTTAA AGAGCGTTAA CGTAGAGCCG GTTTATAAGA GCGAATGAAA ATTGCATAGC

      5770        5780        5790        5800        5810        5820
ATGGACACAG CCGGCATGTA CGAATGCGTG CTACGCAACT ACAGCCACGG CCTCATCATG
TACCTGTGTC GGCCGTACAT GCTTACGCAC GATGCGTTGA TGTCGGTGCC GGAGTAGTAC

      5830        5840        5850        5860        5870        5880
CAACGCTTCG TAATTCTCAC GCAGCTGGAG ACGCTCAGCC GGCCCGACGA ACCTTGCTGC
GTTGCGAAGC ATTAAGAGTG CGTCGACCTC TGCGAGTCGG CCGGGCTGCT TGGAACGACG

      5890        5900        5910        5920        5930        5940
ACACCGGCGT TAGGTCGCTA CTCGCTGGGA GACCAGATCT GGTCGCCGAC GCCCTGGCGT
TGTGGCCGCA ATCCAGCGAT GAGCGACCCT CTGGTCTAGA CCAGCGGCTG CGGGACCGCA

      5950        5960        5970        5980        5990        6000
CTACGGAATC ACGACTGCGG AACGTACCGC GGCTTTCAAC GCAACTACTT CTATATCGGC
GATGCCTTAG TGCTGACGCC TTGCATGGCG CCGAAAGTTG CGTTGATGAA GATATAGCCG

      6010        6020        6030        6040        6050        6060
CGCGCCGACG CCGAGGATTG CTGGAAACCC GCATGTCCGG ACGAGGAACC CGACCGCTGT
GCGCGGCTGC GGCTCCTAAC GACCTTTGGG CGTACAGGCC TGCTCCTTGG GCTGGCGACA

      6070        6080        6090        6100        6110        6120
TGGACAGTGA TACAGCGTTA CCGGCTCCCC GGCGACTGCT ACCGTTCGCA GCCACACCCG
ACCTGTCACT ATGTCGCAAT GGCCGAGGGG CCGCTGACGA TGGCAAGCGT CGGTGTGGGC
```

Fig. 1F

```
      6130        6140        6150        6160        6170        6180
CCGAAATTTT TACCGGTGAC GCCAGCACCG CCGGCCGACA TAGACACCGG GATGTCTCCC
GGCTTTAAAA ATGGCCACTG CGGTCGTGGC GGCCGGCTGT ATCTGTGGCC CTACAGAGGG

      6190        6200        6210        6220        6230        6240
TGGGCCACTC GGGGAATCGC GGCGTTTTTG GGGTTTTGGA GTATTTTTAC CGTATGTTTC
ACCCGGTGAG CCCCTTAGCG CCGCAAAAAC CCCAAAACCT CATAAAAATG GCATACAAAG

      6250        6260        6270        6280        6290        6300
CTATGCTACC TGTGTTATCT GCAGTGTTGT GGACGCTGGT GTCCCACGCC GGGAAGGGGA
GATACGATGG ACACAATAGA CGTCACAACA CCTGCGACCA CAGGGTGCGG CCCTTCCCCT

      6310        6320        6330        6340        6350        6360
CGACGAGGCG GTGAGGGCTA TCGACGCCTA CCGACTTACG ATAGTTACCC CGGTGTTAGA
GCTGCTCCGC CACTCCCGAT AGCTGCGGAT GGCTGAATGC TATCAATGGG GCCACAATCT

      6370        6380        6390        6400        6410        6420
AAGATGAAGA GGTGAGAACA CGTATAAAAT AAAAAAATAA TATGTTAAAA AATGCAGTGT
TTCTACTTCT CCACTCTTGT GCATATTTTA TTTTTTTATT ATACAATTTT TTACGTCACA

      6430        6440        6450        6460        6470        6480
GTGAAGTGTG AATAGTGTGA TTAAAATATG CGGATTGAAT GGGTGTGGTG GTTATTCGGA
CACTTCACAC TTATCACACT AATTTTATAC GCCTAACTTA CCCACACCAC CAATAAGCCT

      6490        6500        6510        6520        6530        6540
TACTTTGTGT CATCCGTTGG GAGCGAACGG TCATTATCCT ATCGTTACCA CTTGGAATCT
ATGAAACACA GTAGGCAACC CTCGCTTGCC AGTAATAGGA TAGCAATGGT GAACCTTAGA

      6550        6560        6570        6580        6590        6600
AATTCATCTA CCAACGTGGT TTGCAACGGA AACATTTCCG TGTTTGTAAA CGGCACCCTA
TTAAGTAGAT GGTTGCACCA AACGTTGCCT TTGTAAAGGC ACAAACATTT GCCGTGGGAT

      6610        6620        6630        6640        6650        6660
GGTGTGCGGT ATAACATTAC GGTAGGAATC AGTTCGTCTT TATTAATAGG ACACCTTACT
CCACACGCCA TATTGTAATG CCATCCTTAG TCAAGCAGAA ATAATTATCC TGTGGAATGA

      6670        6680        6690        6700        6710        6720
ATACAAGTAT TGGAATCATG GTTCACACCC TGGGTCCAAA ATAAAAGTTA CAACAAACAA
TATGTTCATA ACCTTAGTAC CAAGTGTGGG ACCCAGGTTT TATTTTCAAT GTTGTTTGTT

      6730        6740        6750        6760        6770        6780
CCCCTAGGTG ACACTGAAAC GCTTTATAAT ATAGATAGCG AAAACATTCA TCGCGTATCT
GGGGATCCAC TGTGACTTTG CGAAATATTA TATCTATCGC TTTTGTAAGT AGCGCATAGA

      6790        6800        6810        6820        6830        6840
CAATATTTTC ACACAAGATG GATAAAATCT CTGCAAGAGA ATCACACTTG CGACCTCACA
GTTATAAAAG TGTGTTCTAC CTATTTTAGA GACGTTCTCT TAGTGTGAAC GCTGGAGTGT

      6850        6860        6870        6880        6890        6900
AACAGTACAC CTACCTATAC ATATCAAGTA AACGTGAACA ACACGAATTA CCTAACACTA
TTGTCATGTG GATGGATATG TATAGTTCAT TTGCACTTGT TGTGCTTAAT GGATTGTGAT

      6910        6920        6930        6940        6950        6960
ACATCCTCGG GATGGCAAGA CCGTCTAAAT TACACCGTCA TAAATAGTAC ACACTTTAAC
TGTAGGAGCC CTACCGTTCT GGCAGATTTA ATGTGGCAGT ATTTATCATG TGTGAAATTG

      6970        6980        6990        7000        7010        7020
CTCACAGAAT CGAACATAAC CAGCATTCAA AAATATCTCA ACACTACCTG CATAGAAAGA
GAGTGTCTTA GCTTGTATTG GTCGTAAGTT TTTATAGAGT TGTGATGGAC GTATCTTTCT

      7030        7040        7050        7060        7070        7080
CTCCGTAACT ACACCTTGGA GTCCGTATAC ACCACAACTG TGCCTCAAAA CATAACAACA
GAGGCATTGA TGTGGAACCT CAGGCATATG TGGTGTTGAC ACGGAGTTTT GTATTGTTGT

      7090        7100        7110        7120        7130        7140
TCTCAACACG CAACAACCAC TATGCACACA ATACCTCCAA ATACAATAAC AATTCAAAAT
AGAGTTGTGC GTTGTTGGTG ATACGTGTGT TATGGAGGTT TATGTTATTG TTAAGTTTTA
```

Fig. 1G

```
      7150        7160        7170        7180        7190        7200
ACAACTCAAA GCCATACTGT ACAGACGCCG TCTTTTAACG ACACACATAA CGTGACGAAA
TGTTGAGTTT CGGTATGACA TGTCTGCGGC AGAAAATTGC TGTGTGTATT GCACTGCTTT

      7210        7220        7230        7240        7250        7260
CACACGTTAA ACATAAGCTA CGTTTTATCA CAAAAAACGA ATAACACAAC ATCACCGTGG
GTGTGCAATT TGTATTCGAT GCAAAATAGT GTTTTTTGCT TATTGTGTTG TAGTGGCACC

      7270        7280        7290        7300        7310        7320
ATATATGCCA TACCTATGGG CGCTACAGCC ACAATAGGCG CCGGTTTATA TATCGGGAAA
TATATACGGT ATGGATACCC GCGATGTCGG TGTTATCCGC GGCCAAATAT ATAGCCCTTT

      7330        7340        7350        7360        7370        7380
CACTTTACGC CGGTTAAGTT CGTATACGAG GTATGGCGCG GTCAGTAAAG ACGATTCGGA
GTGAAATGCG GCCAATTCAA GCATATGCTC CATACCGCGC CAGTCATTTC TGCTAAGCCT

      7390        7400        7410        7420        7430        7440
TTCAACACAT ATACTCCCCA CGATCCTCGA ACACCTTACA GCATATGAGC AAAAAACAAG
AAGTTGTGTA TATGAGGGGT GCTAGGAGCT TGTGGAATGT CGTATACTCG TTTTTTGTTC

      7450        7460        7470        7480        7490        7500
AAAGTATAGC CACAATCACA TTTGGGCGAA TAACATGCTG TCATCCACTA GCGTCTATTA
TTTCATATCG GTGTTAGTGT AAACCCGCTT ATTGTACGAC AGTAGGTGAT CGCAGATAAT

      7510        7520        7530        7540        7550        7560
ATCTAATGTT TAACGGGAGC TGTACTGTCA CCGTTAAAAT ATCCATGGGA ATCAACGGGT
TAGATTACAA ATTGCCCTCG ACATGACAGT GGCAATTTTA TAGGTACCCT TAGTTGCCCA

      7570        7580        7590        7600        7610        7620
CAACCAACGT CCATCAGCTT GTGATTGTGC TCCATCTGGG TAACCGCTGT CAGCCTTGGC
GTTGGTTGCA GGTAGTCGAA CACTAACACG AGGTAGACCC ATTGGCGACA GTCGGAACCG

      7630        7640        7650        7660        7670        7680
GACAGGTGTA ATCACAGCTG TCACATAACT CACGAAGCCT CCAATCACAG CAGCACACAT
CTGTCCACAT TAGTGTCGAC AGTGTATTGA GTGCTTCGGA GGTTAGTGTC GTCGTGTGTA

      7690        7700        7710        7720        7730        7740
AGTCCTAACG CCATTGGCGT GTATAAAAGT TCGGAAAACT TGACGGTTGT ACGGCACGAC
TCAGGATTGC GGTAACCGCA CATATTTTCA AGCCTTTTGA ACTGCCAACA TGCCGTGCTG

      7750        7760        7770        7780        7790        7800
AAATCGATGT AGTGGTATGT TTTTCCAGCA GAGACCGTGT GCGGTCTCTT AGGTTCGCTA
TTTAGCTACA TCACCATACA AAAAGGTCGT CTCTGGCACA CGCCAGAGAA TCCAAGCGAT

      7810        7820        7830        7840        7850        7860
TACTGTGGCT GGAAACTGGT TACCTGTGAA GATGGCTAAC TATCCTGTTC TGTCCTGGAA
ATGACACCGA CCTTTGACCA ATGGACACTT CTACCGATTG ATAGGACAAG ACAGGACCTT

      7870        7880        7890        7900        7910        7920
AAACTTTTGG CGTCGTAGGT GGACTTTGCA GTATGCGGGT TAGTGAAGTT ATGTCATTTA
TTTGAAAACC GCAGCATCCA CCTGAAACGT CATACGCCCA ATCACTTCAA TACAGTAAAT

      7930        7940        7950        7960        7970        7980
TTTACGTTTA CGATCTCGTA TTACAAACCG CGGAGAGGAT GATACCGTTC GGCCCCATGA
AAATGCAAAT GCTAGAGCAT AATGTTTGGC GCCTCTCCTA CTATGGCAAG CCGGGGTACT

      7990        8000        8010        8020        8030        8040
GTTATTTTTA TTCTTCCGGT AGGAGGCATG AAGCCTCTGA TAATGCTCAT CTGCTTTGCT
CAATAAAAAT AAGAAGGCCA TCCTCCGTAC TTCGGAGACT ATTACGAGTA GACGAAACGA

      8050        8060        8070        8080        8090        8100
GTGATATTAT TGCAGCTTGG AGTGACTAAA GTGTGTCAGC ATAATGAAGT GCAACTGGGC
CACTATAATA ACGTCGAACC TCACTGATTT CACACAGTCG TATTACTTCA CGTTGACCCG

      8110        8120        8130        8140        8150        8160
AATGAGTGCT GCCCTCCGTG TGGTTCGGGA CAAAGAGTTA CTAAAGTATG CACGGATTAT
TTACTCACGA CGGGAGGCAC ACCAAGCCCT GTTTCTCAAT GATTTCATAC GTGCCTAATA
```

Fig. 1H

```
      8170        8180        8190        8200        8210        8220
ACCAGTGTAA CGTGTACCCC TTGCCCCAAC GGCACGTATG TATCGGGACT TTACAACTGT
TGGTCACATT GCACATGGGG AACGGGGTTG CCGTGCATAC ATAGCCCTGA AATGTTGACA

      8230        8240        8250        8260        8270        8280
ACCGATTGCA CTCAATGTAA CGTCACTCAG GTCATGATTC GTAACTGCAC TTCCACCAAT
TGGCTAACGT GAGTTACATT GCAGTGAGTC CAGTACTAAG CATTGACGTG AAGGTGGTTA

      8290        8300        8310        8320        8330        8340
AATACCGTAT GCGCACCTAA GAACCATACG TACTTTTCCA CTCCAGGCGT CCAACATCAC
TTATGGCATA CGCGTGGATT CTTGGTATGC ATGAAAAGGT GAGGTCCGCA GGTTGTAGTG

      8350        8360        8370        8380        8390        8400
AAACAACGAC AGCAAAATCA TACCGCACAT ATAACCGTCA AACAAGGAAA AAGCGGTCGT
TTTGTTGCTG TCGTTTTAGT ATGGCGTGTA TATTGGCAGT TTGTTCCTTT TTCGCCAGCA

      8410        8420        8430        8440        8450        8460
CATACTCTAG CCTGGTTGTC TCTCTTTATC TTTCTTGTGG GTATCATACT TTTAATTCTC
GTATGAGATC GGACCAACAG AGAGAAATAG AAAGAACACC CATAGTATGA AAATTAAGAG

      8470        8480        8490        8500        8510        8520
TATCTTATAG CCGCCTATCG GAGTGAGAGA TGCCAACAGT GTTGCTCAAT CGGCAAAATT
ATAGAATATC GGCGGATAGC CTCACTCTCT ACGGTTGTCA CAACGAGTTA GCCGTTTTAA

      8530        8540        8550        8560        8570        8580
TTCTACCGCA CCCTGTAAGC TTCCTGTTGT TGTTTTTACA TCACGGTACG ATGAAGTCAC
AAGATGGCGT GGGACATTCG AAGGACAACA ACAAAAATGT AGTGCCATGC TACTTCAGTG

      8590        8600        8610        8620        8630        8640
ACAGATAATT ACAGATGAGC TGTTCATATT TTTTATTATT TTTTCCAATT CCTGCACTAA
TGTCTATTAA TGTCTACTCG ACAAGTATAA AAAATAATAA AAAAGGTTAA GGACGTGATT

      8650        8660        8670        8680        8690        8700
AAAAAGAAGC ACTTTACGGA ACCGTGTCTG AGTATCTGTG GGGAATTTAG GTACTTTTTG
TTTTTCTTCG TGAAATGCCT TGGCACAGAC TCATAGACAC CCCTTAAATC CATGAAAAAC

      8710        8720        8730        8740        8750        8760
CCGACGTCAG GAAAAATAAG TGTCGCCTAC ATAAGAGCCC GGTGCTATCG TGCTGTCACT
GGCTGCAGTC CTTTTTATTC ACAGCGGATG TATTCTCGGG CCACGATAGC ACGACAGTGA

      8770        8780        8790        8800        8810        8820
CTTTCTTGTT GCCTTCGATG TACGGCGTCC TGGCTCATTA CTACTCCTTC ATCAGTAGCC
GAAAGAACAA CGGAAGCTAC ATGCCGCAGG ACCGAGTAAT GATGAGGAAG TAGTCATCGG

      8830        8840        8850        8860        8870        8880
CCAGCGTTAT GGTTAATTTT AAGCATCATA ACGCCGTGCA GCTGTTATGT GCACGGACCC
GGTCGCAATA CCAATTAAAA TTCGTAGTAT TGCGGCACGT CGACAATACA CGTGCCTGGG

      8890        8900        8910        8920        8930        8940
GAGACGCACT GCCGGATGGG AACGTTTAAC CCATCATGCG TCGTATCACG CGAACTACGG
CTCTGCGTGA CGGCCTACCC TTGCAAATTG GGTAGTACGC AGCATAGTGC GCTTGATGCC

      8950        8960        8970        8980        8990        9000
GGCATACGCC GTGTTGATGG CTACATCGCA AAGAAAGTCC CTAGTGTTAC ATCGATACAG
CCGTATGCGG CACAACTACC GATGTAGCGT TTCTTTCAGG GATCACAATG TAGCTATGTC

      9010        9020        9030        9040        9050        9060
TGCCGTGACA GCCGTGGCCC TGCAGCTCAT GCCTGTTGAG ATCGTCCGCA AGCTAGATCA
ACGGCACTGT CGGCACCGGG ACGTCGAGTA CGGACAACTC TAGCAGGCGT TCGATCTAGT

      9070        9080        9090        9100        9110        9120
GTCGGACTGG GTGCGGGGTG CCTGGATCGT GTCAGAGACT TTTCCAACTA GCGACCCCAA
CAGCCTGACC CACGCCCCAC GGACCTAGCA CAGTCTCTGA AAAGGTTGAT CGCTGGGGTT

      9130        9140        9150        9160        9170        9180
AGGAGTTTGG AGCGACGATG ACTCCTCGAT GGGTGGAAGT GATGATTGAT GATGAGAACC
TCCTCAAACC TCGCTGCTAC TGAGGAGCTA CCCACCTTCA CTACTAACTA CTACTCTTGG
```

Fig. 1I

```
      9190        9200        9210        9220        9230        9240
TGACAAGAAA GACGAGAGAG AAATTTAGAG CTGTCATTGT AGAATTAGTC TAGATTCCTG
ACTGTTCTTT CTGCTCTCTC TTTAAATCTC GACAGTAACA TCTTAATCAG ATCTAAGGAC

      9250        9260        9270        9280        9290        9300
ATAATAAACA GTATCGATTT TGAAACCTAA TTGACGTGTG ATCGATTTTT AAACCTCTGT
TATTATTTGT CATAGCTAAA ACTTTGGATT AACTGCACAC TAGCTAAAAA TTTGGAGACA

      9310        9320        9330        9340        9350        9360
GTTGTGTGAT TGATTGGTAT GTGGGGGGAT CCGATTTCAA AGGGGGGTAC TTATCGGGAA
CAACACACTA ACTAACCATA CACCCCCCTA GGCTAAAGTT TCCCCCCATG AATAGCCCTT

      9370        9380        9390        9400        9410        9420
TTGATGTGTC ATGGACGCAG TTTTGAGCGA TTTTCCGGGA ATACCGGATA TTACGAATTA
AACTACACAG TACCTGCGTC AAAACTCGCT AAAAGGCCCT TATGGCCTAT AATGCTTAAT

      9430        9440        9450        9460        9470        9480
CTGGTAGTGA CGTAGATAAT AAAATTATAA TGCGATTAAT TTTTGGTGCG TTGATTATTT
GACCATCACT GCATCTATTA TTTTAATATT ACGCTAATTA AAAACCACGC AACTAATAAA

      9490        9500        9510        9520        9530        9540
TTTTAGCATA TGTGTATCAT TATGAGGTGA ATGGAACAGA ATTACGCTGC AGATGTCTTC
AAAATCGTAT ACACATAGTA ATACTCCACT TACCTTGTCT TAATGCGACG TCTACAGAAG

      9550        9560        9570        9580        9590        9600
ATAGAAAATG GCCGCCTAAT AAAATTATAT TGGGTAATTA TTGGCTTCAT CGCGATCCCA
TATCTTTTAC CGGCGGATTA TTTTAATATA ACCCATTAAT AACCGAAGTA GCGCTAGGGT

      9610        9620        9630        9640        9650        9660
GAGGGCCCGG ATGCGATAAA AATGAACATT TATTGTATCC AGACGGAAGG AAACCGCCTG
CTCCCGGGCC TACGCTATTT TTACTTGTAA ATAACATAGG TCTGCCTTCC TTTGGCGGAC

      9670        9680        9690        9700        9710        9720
GACCTGGAGT ATGTTTATCG CCCGATCACC TCTTCTCAAA ATGGTTAGAC AAACACAACG
CTGGACCTCA TACAAATAGC GGGCTAGTGG AGAAGAGTTT TACCAATCTG TTTGTGTTGC

      9730        9740        9750        9760        9770        9780
ATAATAGGTG GTATAATGTT AACATAACGA AATCACCAGG ACCGAGACGA ATAAATATAA
TATTATCCAC CATATTACAA TTGTATTGCT TTAGTGGTCC TGGCTCTGCT TATTTATATT

      9790        9800        9810        9820        9830        9840
CCTTGATAGG TGTTAGAGGA TAATATTTAA TGTATGTTTT CAAACAGACA AGTTCGTTAA
GGAACTATCC ACAATCTCCT ATTATAAATT ACATACAAAA GTTTGTCTGT TCAAGCAATT

      9850        9860        9870        9880        9890        9900
AACAAAATAT TACAGTATGT GTTTAATATG GTGCTAACAT GGTTGCACCA TCCGGTTTCA
TTGTTTTATA ATGTCATACA CAAATTATAC CACGATTGTA CCAACGTGGT AGGCCAAAGT

      9910        9920        9930        9940        9950        9960
AACTCGCATA TCAATCTGTT ATCGGTACGA CACCTGTCAT TAATCGCATA TATGTTACTT
TTGAGCGTAT AGTTAGACAA TAGCCATGCT GTGGACAGTA ATTAGCGTAT ATACAATGAA

      9970        9980        9990       10000       10010       10020
ACCATATGTC CCCTAGCCGT CCATGTTTTA GAACTAGAAG ATTACGACAG GCGCTGCCGT
TGGTATACAG GGGATCGGCA GGTACAAAAT CTTGATCTTC TAATGCTGTC CGCGACGGCA

      10030       10040       10050       10060       10070       10080
TGCAACAACC AAATTCTGTT GAATACCCTG CCGGTCGGAA CCGAATTGCT TAAGCCAATC
ACGTTGTTGG TTTAAGACAA CTTATGGGAC GGCCAGCCTT GGCTTAACGA ATTCGGTTAG

      10090       10100       10110       10120       10130       10140
GCAGCGAGCG AAAGCTGCAA TCGTCAGGAA GTGCTGGCTA TTTTAAAGGA CAAGGGAACC
CGTCGCTCGC TTTCGACGTT AGCAGTCCTT CACGACCGAT AAAATTTCCT GTTCCCTTGG

      10150       10160       10170       10180       10190       10200
AAGTGTCTCA ATCCTAACGC GCAAGCCGTG CGTCGTCACA TCAACCGGCT ATTTTTTCGG
TTCACAGAGT TAGGATTGCG CGTTCGGCAC GCAGCAGTGT AGTTGGCCGA TAAAAAAGCC
```

Fig. 1J

```
     10210      10220      10230      10240      10250      10260
TTAATCTTAG ACGAGGAACA [illegible]CATTTAC GACGTAGTGT CTACCAATAT [illegible]AGTTCGGT
AATTAGAATC TGCTCCTTGT [illegible]GTAAATG CTGCATCACA GATGGTTATA [illegible]TCAAGCCA

     10270      10280      10290      10300      10310      10320
GCCTGGCCAG TCCCTACGGC CTACAAAGCC TTTCTTTGGA AATACGCCAA GAGACTGAAC
CGGACCGGTC AGGGATGCCG GATGTTTCGG AAAGAAACCT TTATGCGGTT CTCTGACTTG

     10330      10340      10350      10360      10370      10380
TACCACCACT TCAGACTGCG CTGGTGATCA TGTCCCTATT TTACCGTGCG GTAGCTCTGG
ATGGTGGTGA AGTCTGACGC GACCACTAGT ACAGGGATAA AATGGCACGC CATCGAGACC

     10390      10400      10410      10420      10430      10440
GCACGCTAAG CGCTTTGGTG TGGTACAGCA CTAGCATCCT CGCAGAGATT AACGAAAATT
CGTGCGATTC GCGAAACCAC ACCATGTCGT GATCGTAGGA GCGTCTCTAA TTGCTTTTAA

     10450      10460      10470      10480      10490      10500
CCTGCTCCTC ATCTTCTGCG GATCACGAAG ACTGCGAGGA ACCGGACGAG ATCGTTCGCG
GGACGAGGAG TAGAAGACGC CTAGTGCTTC TGACGCTCCT TGGCCTGCTC TAGCAAGCGC

     10510      10520      10530      10540      10550      10560
AAGAGCAAGA CTATCGGGCT CTGCTGGCCT TTTCCCTAGT GATTTGCGGT ACGCTCCTCG
TTCTCGTTCT GATAGCCCGA GACGACCGGA AAAGGGATCA CTAAACGCCA TGCGAGGAGC

     10570      10580      10590      10600      10610      10620
TCACTTGTGT GATCTGAGAC GTCATGCTGG TAGCGTTTAT GAGTCGGGCG GTGGCCGACA
AGTGAACACA CTAGACTCTG CAGTACGACC ATCGCAAATA CTCAGCCCGC CACCGGCTGT

     10630      10640      10650      10660      10670      10680
CGCCGCATTT CCTAACCCGC GCAGCATGTT GCGCTTGCTG TTCACGCTCG TCCTGCTGGC
GCGGCGTAAA GGATTGGGCG CGTCGTACAA CGCGAACGAC AAGTGCGAGC AGGACGACCG

     10690      10700      10710      10720      10730      10740
CCTCCACGGG CAGTCTGTCG GCGCTAGCCG CGACTATGTG CATGTTCGGC TACTGAGCTA
GGAGGTGCCC GTCAGACAGC CGCGATCGGC GCTGATACAC GTACAAGCCG ATGACTCGAT

     10750      10760      10770      10780      10790      10800
CCGAGGCGAC CCCCTGGTCT TCAAGCACAC TTTCTCGGGT GTGCGTCGAC CCTTCACCGA
GGCTCCGCTG GGGGACCAGA AGTTCGTGTG AAAGAGCCCA CACGCAGCTG GGAAGTGGCT

     10810      10820      10830      10840      10850      10860
GCTAGGCTGG GCTGCGTGTC GCGACTGGGA CAGTATGCAT TGCACACCCT TCTGGTCTAC
CGATCCGACC CGACGCACAG CGCTGACCCT GTCATACGTA ACGTGTGGGA AGACCAGATG

     10870      10880      10890      10900      10910      10920
CGATCTGGAG CAGATGACCG ACTCGGTGCG GCGTTACAGC ACGGTGAGCC CCGGCAAGGA
GCTAGACCTC GTCTACTGGC TGAGCCACGC CGCAATGTCG TGCCACTCGG GGCCGTTCCT

     10930      10940      10950      10960      10970      10980
AGTGACGCTT CAGCTTCACG GGAACCAAAC CGTACAGCCG TCGTTTCTAA GCTTTACGTG
TCACTGCGAA GTCGAAGTGC CCTTGGTTTG GCATGTCGGC AGCAAAGATT CGAAATGCAC

     10990      11000      11010      11020      11030      11040
CCGCCTGCAG CTAGAACCCG TGGTGGAAAA TGTTGGCCTC TACGTGGCCT ACGTGGTCAA
GGCGGACGTC GATCTTGGGC ACCACCTTTT ACAACCGGAG ATGCACCGGA TGCACCAGTT

     11050      11060      11070      11080      11090      11100
CGACGGCGAA CGCCCACAAC AGTTTTTTAC ACCGCAGGTA GACGTGGTAC GCTTTGCTCT
GCTGCCGCTT GCGGGTGTTG TCAAAAAATG TGGCGTCCAT CTGCACCATG CGAAACGAGA

     11110      11120      11130      11140      11150      11160
ATATCTAGAA ACACTCTCCC GGATCGTGGA ACCGTTAGAA TCAGGTCGCC TGGCAGTGGA
TATAGATCTT TGTGAGAGGG CCTAGCACCT TGGCAATCTT AGTCCAGCGG ACCGTCACCT

     11170      11180      11190      11200      11210      11220
ATTTGATACG CCTGACCTAG CTCTGGCGCC CGATTTAGTA AGCAGCCTCT TCGTGGCCGG
TAAACTATGC GGACTGGATC GAGACCGCGG GCTAAATCAT TCGTCGGAGA AGCACCGGCC
```

Fig. 1K

```
      11230       11240       11250       11260       11270       11280
ACACGGCGAG ACCGACTTTT ACATGAACTG GACGCTGCGT CGCAGTCAGA CCCACTACCT
TGTGCCGCTC TGGCTGAAAA TGTACTTGAC CTGCGACGCA GCGTCAGTCT GGGTGATGGA

      11290       11300       11310       11320       11330       11340
GGAGGAGATG GCCTTACAGG TGGAGATTCT AAAACCCCGC GGCGTACGTC ACCGCGCTAT
CCTCCTCTAC CGGAATGTCC ACCTCTAAGA TTTTGGGGCG CCGCATGCAG TGGCGCGATA

      11350       11360       11370       11380       11390       11400
TATCCACCAT CCGAAGCTAC AGCCGGGCGT TGGCCTGTGG ATAGATTTCT GCGTGTACCG
ATAGGTGGTA GGCTTCGATG TCGGCCCGCA ACCGGACACC TATCTAAAGA CGCACATGGC

      11410       11420       11430       11440       11450       11460
CTACAACGCG CGCCTGACCC GCGGCTACGT ACGATACACC CTGTCACCGA AAGCGCGCTT
GATGTTGCGC GCGGACTGGG CGCCGATGCA TGCTATGTGG GACAGTGGCT TTCGCGCGAA

      11470       11480       11490       11500       11510       11520
GCCCGCAAAA GCAGAGGGTT GGCTGGTGTC ACTAGACAGA TTCATCGTGC AGTACCTCAA
CGGGCGTTTT CGTCTCCCAA CCGACCACAG TGATCTGTCT AAGTAGCACG TCATGGAGTT

      11530       11540       11550       11560       11570       11580
CACATTGCTG ATTACAATGA TGGCGGCGAT ATGGGCTCGC GTTTTGATAA CCTACCTGGT
GTGTAACGAC TAATGTTACT ACCGCCGCTA TACCCGAGCG CAAAACTATT GGATGGACCA

      11590       11600       11610       11620       11630       11640
GTCGCGGCGT CGGTAGAGGC TTGCGGAAAC CACGTCCTCG TCACACGTCG TTCGCGGACA
CAGCGCCGCA GCCATCTCCG AACGCCTTTG GTGCAGGAGC AGTGTGCAGC AAGCGCCTGT

      11650       11660       11670       11680       11690       11700
TAGCAAGAAA TCCACGTCGC CACATCTCGA GAATGCCGGC CTTGCGGGGT CCCCTTCGCG
ATCGTTCTTT AGGTGCAGCG GTGTAGAGCT CTTACGGCCG GAACGCCCCA GGGGAAGCGC

      11710       11720       11730       11740       11750       11760
CAACATTCCT GGCCCTGGTC GCGTTCGGGT TGCTGCTTCA GATAGACCTC AGCGACGCTA
GTTGTAAGGA CCGGGACCAG CGCAAGCCCA ACGACGAAGT CTATCTGGAG TCGCTGCGAT

      11770       11780       11790       11800       11810       11820
CGAATGTGAC CAGCAGCACA AAAGTCCCTA CTAGCACCAG CAACAGAAAT AACGTCGACA
GCTTACACTG GTCGTCGTGT TTTCAGGGAT GATCGTGGTC GTTGTCTTTA TTGCAGCTGT

      11830       11840       11850       11860       11870       11880
ACGCCACGAG TAGCGGACCC ACAACCGGGA TCAACATGAC CACCACCCAC GAGTCTTCCG
TGCGGTGCTC ATCGCCTGGG TGTTGGCCCT AGTTGTACTG GTGGTGGGTG CTCAGAAGGC

      11890       11900       11910       11920       11930       11940
TTCACAACGT GCGCAATAAC GAGATCATGA AAGTGCTGGC TATCCTCTTC TACATCGTGA
AAGTGTTGCA CGCGTTATTG CTCTAGTACT TTCACGACCG ATAGGAGAAG ATGTAGCACT

      11950       11960       11970       11980       11990       12000
CAGGCACCTC CATTTTCAGC TTCATAGCGG TACTGATCGC GGTAGTTTAC TCCTCGTGTT
GTCCGTGGAG GTAAAAGTCG AAGTATCGCC ATGACTAGCG CCATCAAATG AGGAGCACAA

      12010       12020       12030       12040       12050       12060
GCAAGCACCC GGGCCGCTTT CGTTTCGCCG ACGAAGAGGC CGTCAACCTG TTGGACGACA
CGTTCGTGGG CCCGGCGAAA GCAAAGCGGC TGCTTCTCCG GCAGTTGGAC AACCTGCTGT

      12070       12080       12090       12100       12110       12120
CGGACGACAG TGGCGGCAGC AGCCCGTTTG GCAGCGGTTC CCGACGAGGT TCTCAGATCC
GCCTGCTGTC ACCGCCGTCG TCGGGCAAAC CGTCGCCAAG GGCTGCTCCA AGAGTCTAGG

      12130       12140       12150       12160       12170       12180
CCGCCGGATT TTGTTCCTCG AGCCCTTATC AGCGGTTGGA AACTCGGGAC TGGGACGAGG
GGCGGCCTAA AACAAGGAGC TCGGGAATAG TCGCCAACCT TTGAGCCCTG ACCCTGCTCC

      12190       12200       12210       12220       12230       12240
AGGAGGAGGC GTCCGCGGCC CGCGAGCGCA TGAAACATGA TCCTGAGAAC GTCATCTATT
TCCTCCTCCG CAGGCGCCGG GCGCTCGCGT ACTTTGTACT AGGACTCTTG CAGTAGATAA
```

Fig. 1L

```
     12250       12260       12270       12280       12290       12300
TCAGAAAGGA TGGCAACTTG GACACGTCGT TCGTGAATCC CAATTATGGG AGAGGCTCGC
AGTCTTTCCT ACCGTTGAAC CTGTGCAGCA AGCACTTAGG GTTAATACCC TCTCCGAGCG

     12310       12320       12330       12340       12350       12360
CTTTGACCAT CGAATCTCAC CTCTCGGACA ATGAGGAGGA CCCCATCAGG TACTACGTTT
GAAACTGGTA GCTTAGAGTG GAGAGCCTGT TACTCCTCCT GGGGTAGTCC ATGATGCAAA

     12370       12380       12390       12400       12410       12420
CGGTGTACGA TGAACTGACC GCCTCGGAAA TGGAAGAACC TTCGAACAGC ACCAGCTGGC
GCCACATGCT ACTTGACTGG CGGAGCCTTT ACCTTCTTGG AAGCTTGTCG TGGTCGACCG

     12430       12440       12450       12460       12470       12480
AGATTCCCAA ACTAATGAAA GTTGCCATGC AACCCGTCTC GCTCAGAGAT CCCGAGTACG
TCTAAGGGTT TGATTACTTT CAACGGTACG TTGGGCAGAG CGAGTCTCTA GGGCTCATGC

     12490       12500       12510       12520       12530       12540
ACTAGGCTTT TTTTTTTGTC TTTCGGTTCC AACTCTTTCC CCGCCCCATC ACCTCGCCTG
TGATCCGAAA AAAAAAACAG AAAGCCAAGG TTGAGAAAGG GGCGGGGTAG TGGAGCGGAC

     12550       12560       12570       12580       12590       12600
TACTATGTGT ATGATGTCTC ATAATAAAGC TTTCTTTCTC AGTCTGCAAC ATGCAGCTGT
ATGATACACA TACTACAGAG TATTATTTCG AAAGAAAGAG TCAGACGTTG TACGTCGACA

     12610       12620       12630       12640       12650       12660
GTCGGGTGTG GCTGTCTGTT TGTCTGTGCG CCGTGGTGCT GGGTCAGTGC CAGCGGGAAA
CAGCCCACAC CGACAGACAA ACAGACACGC GGCACCACGA CCCAGTCACG GTCGCCCTTT

     12670       12680       12690       12700       12710       12720
CCGCGGAAAA AAACGATTAT TACCGAGTAC CGCATTACTG GGACGCGTGC TCTCGCGCGC
GGCGCCTTTT TTTGCTAATA ATGGCTCATG GCGTAATGAC CCTGCGCACG AGAGCGCGCG

     12730       12740       12750       12760       12770       12780
TGCCCGACCA AACCCGTTAC AAGTATGTGG AACAGCTCGT GGACCTCACG TTGAACTACC
ACGGGCTGGT TTGGGCAATG TTCATACACC TTGTCGAGCA CCTGGAGTGC AACTTGATGG

     12790       12800       12810       12820       12830       12840
ACTACGATGC GAGCCACGGC TTGGACAACT TTGACGTGCT CAAGAGGTGA GGGTACGCGC
TGATGCTACG CTCGGTGCCG AACCTGTTGA AACTGCACGA GTTCTCCACT CCCATGCGCG

     12850       12860       12870       12880       12890       12900
TAAAGGTGCA TGACAACGGG AAGGTAAGGG CGAACGGGTA ACGGCTAAGT AACCGCATGG
ATTTCCACGT ACTGTTGCCC TTCCATTCCC GCTTGCCCAT TGCCGATTCA TTGGCGTACC

     12910       12920       12930       12940       12950       12960
GGTATGAAAT GACGTTTGGA ACCTGTGCTT GCAGAATCAA CGTGACCGAG GTGTCGTTGC
CCATACTTTA CTGCAAACCT TGGACACGAA CGTCTTAGTT GCACTGGCTC CACAGCAACG

     12970       12980       12990       13000       13010       13020
TCATCAGCGA CTTTAGACGT CAGAACCGTC GCGGCGGCAC CAACAAAAGG ACCACGTTCA
AGTAGTCGCT GAAATCTGCA GTCTTGGCAG CGCCGCCGTG GTTGTTTTCC TGGTGCAAGT

     13030       13040       13050       13060       13070       13080
ACGCCGCCGG TTCGCTGGCG CCACACGCCC GGAGCCTCGA GTTCAGCGTG CGGCTCTTTG
TGCGGCGGCC AAGCGACCGC GGTGTGCGGG CCTCGGAGCT CAAGTCGCAC GCCGAGAAAC

     13090       13100       13110       13120       13130       13140
CCAACTAGCC TGCGTCACGG GAAATAATAT GCTGCGGCTT CTGCTTCGTC ACCACTTTCA
GGTTGATCGG ACGCAGTGCC CTTTATTATA CGACGCCGAA GACGAAGCAG TGGTGAAAGT

     13150       13160       13170       13180       13190       13200
CTGCCTGCTT CTGTGCGCGG TTTGGGCAAC GCCCTGTCTG GCGTCTCCGT GGTCGACGCT
GACGGACGAA GACACGCGCC AAACCCGTTG CGGGACAGAC CGCAGAGGCA CCAGCTGCGA

     13210       13220       13230       13240       13250       13260
AACGGCAAAC CAGAATCCGT CCCCGCCATG GTCTAAACTG ACGTATTCCA AACCGCATGA
TTGCCGTTTG GTCTTAGGCA GGGGCGGTAC CAGATTTGAC TGCATAAGGT TTGGCGTACT
```

Fig. 1M

```
      13270       13280       13290       13300       13310       13320
CGCGGCGACG TTTTACTGTC CTTTTCTCTA TCCCTCGCCC CCACGGTCCC CCTTGCAATT
GCGCCGCTGC AAAATGACAG GAAAAGAGAT AGGGAGCGGG GGTGCCAGGG GGAACGTTAA

      13330       13340       13350       13360       13370       13380
CTCGGGGTTC CAGCAGGTAT CAACGGGTCC CGAGTGTCGC AACGAGACCC TGTATCTGCT
GAGCCCCAAG GTCGTCCATA GTTGCCCAGG GCTCACAGCG TTGCTCTGGG ACATAGACGA

      13390       13400       13410       13420       13430       13440
GTACAACCGG GAAGGCCAGA CCTTGGTGGA GAGAAGCTCC ACCTGGGTGA AAAAGGTGAT
CATGTTGGCC CTTCCGGTCT GGAACCACCT CTCTTCGAGG TGGACCCACT TTTTCCACTA

      13450       13460       13470       13480       13490       13500
CTGGTATCTG AGCGGTCGCA ACCAGACCAT CCTCCAACGG ATGCCCCAAA CGGCTTCGAA
GACCATAGAC TCGCCAGCGT TGGTCTGGTA GGAGGTTGCC TACGGGGTTT GCCGAAGCTT

      13510       13520       13530       13540       13550       13560
ACCGAGCGAC GGAAACGTGC AGATCAGCGT GGAAGACGCC AAGATTTTTG GAGCGCACAT
TGGCTCGCTG CCTTTGCACG TCTAGTCGCA CCTTCTGCGG TTCTAAAAAC CTCGCGTGTA

      13570       13580       13590       13600       13610       13620
GGTGCCCAAG CAGACCAAGC TGCTACGCTT CGTCGTCAAC GATGGCACGC GTTATCAGAT
CCACGGGTTC GTCTGGTTCG ACGATGCGAA GCAGCAGTTG CTACCGTGCG CAATAGTCTA

      13630       13640       13650       13660       13670       13680
GTGTGTGATG AAGCTGGAGA GCTGGGCCCA CGTCTTCCGG GACTACAGCG TGTCTTTTCA
CACACACTAC TTCGACCTCT CGACCCGGGT GCAGAAGGCC CTGATGTCGC ACAGAAAAGT

      13690       13700       13710       13720       13730       13740
GGTGCGATTG ACGTTCACCG AGGCCAATAA CCAGACTTAC ACCTTCTGTA CCCATCCCAA
CCACGCTAAC TGCAAGTGGC TCCGGTTATT GGTCTGAATG TGGAAGACAT GGGTAGGGTT

      13750       13760       13770       13780       13790       13800
TCTCATCATT TGAGCCCGTC GCGCGCGCAG GGAATTTTGA AAACCGCGCG TCATGAGTCC
AGAGTAGTAA ACTCGGGCAG CGCGCGCGTC CCTTAAAACT TTTGGCGCGC AGTACTCAGG

      13810       13820       13830       13840       13850       13860
CAAAGACCTG ACGCCGTTCT TGACGACGTT GTGGCTGCTA TTGGGTCACA GCCGCGTGCC
GTTTCTGGAC TGCGGCAAGA ACTGCTGCAA CACCGACGAT AACCCAGTGT CGGCGCACGG

      13870       13880       13890       13900       13910       13920
GCGGGTGCGC GCAGAAGAAT GTTGCGAATT CATAAACGTC AACCACCCGC CGGAACGCTG
CGCCCACGCG CGTCTTCTTA CAACGCTTAA GTATTTGCAG TTGGTGGGCG GCCTTGCGAC

      13930       13940       13950       13960       13970       13980
TTACGATTTC AAAATGTGCA ATCGCTTCAC CGTCGCGTAC GTATTTTCAT GATTGTCTGC
AATGCTAAAG TTTTACACGT TAGCGAAGTG GCAGCGCATG CATAAAAGTA CTAACAGACG

      13990       14000       14010       14020       14030       14040
GTTCTGTGGT GCGTCTGGAT TTGTCTCTCG ACGTTTCTGA TAGCCATGTT CCATCGACGA
CAAGACACCA CGCAGACCTA AACAGAGAGC TGCAAAGACT ATCGGTACAA GGTAGCTGCT

      14050       14060       14070       14080       14090       14100
TCCTCGGGAA TGCCAGAGTA GATTTTCATG AATCCACAGG CTGCGGTGTC CGGACGGCGA
AGGAGCCCTT ACGGTCTCAT CTAAAAGTAC TTAGGTGTCC GACGCCACAG GCCTGCCGCT

      14110       14120       14130       14140       14150       14160
AGTCTGCTAC AGTCCCGAGA AAACGGCTGA GATTCGCGGG ATCGTCACCA CCATGACCCA
TCAGACGATG TCAGGGCTCT TTTGCCGACT CTAAGCGCCC TAGCAGTGGT GGTACTGGGT

      14170       14180       14190       14200       14210       14220
TTCATTGACA CGCCAGGTCG TACACAACAA ACTGACGAGC TGCAACTACA ATCCGTAAGT
AAGTAACTGT GCGGTCCAGC ATGTGTTGTT TGACTGCTCG ACGTTGATGT TAGGCATTCA

      14230       14240       14250       14260       14270       14280
CTCTTCCTCG AGGGCCTTAC AGCCTATGGG AGAGTAAGAC AGAGAGGGAC AAAACATCAT
GAGAAGGAGC TCCCGGAATG TCGGATACCC TCTCATTCTG TCTCTCCCTG TTTTGTAGTA
```

Fig. 1N

```
     14290      14300      14310      14320      14330      14340
TAAAAAAAAA AGTCTAATTT CACGTTTTGT ACCCCCCTTC CCCTCCGTGT TGTAGCCCAT
ATTTTTTTTT TCAGATTAAA GTGCAAAACA TGGGGGGAAG GGGAGGCACA ACATCGGGTA

     14350      14360      14370      14380      14390      14400
CGGCCGCGGC GATCTCCTAG TAACACTCGT CCGACACTTC CACCATCTCC AGCTCGGCCG
GCCGGCGCCG CTAGAGGATC ATTGTGAGCA GGCTGTGAAG GTGGTAGAGG TCGAGCCGGC

     14410      14420      14430      14440      14450      14460
GCGGTTCGGC ATCCTCTACC AGCGGCGTCG TCTCATCTTT GCCGCAGCAG CGGACGCACA
CGCCAAGCCG TAGGAGATGG TCGCCGCAGC AGAGTAGAAA CGGCGTCGTC GCCTGCGTGT

     14470      14480      14490      14500      14510      14520
CCTTCTCCAG GCAGAACGCC ACCAGCTGCC GCCGAACGTA CCACAGGTAC ACGTGCAGAC
GGAAGAGGTC CGTCTTGCGG TGGTCGACGG CGGCTTGCAT GGTGTCCATG TGCACGTCTG

     14530      14540      14550      14560      14570      14580
CTGCGAACAG GACTACGGAG GTCATGACCA CCACGACGCA CACGGGAATC CAGGGATCGA
GACGCTTGTC CTGATGCCTC CAGTACTGGT GGTGCTGCGT GTGCCCTTAG GTCCCTAGCT

     14590      14600      14610      14620      14630      14640
GATTGTTGCT GGAACTCATG GCTATCGCCA CCGACGTGCC CGCGTCTGTC TCACCGCCGC
CTAACAACGA CCTTGAGTAC CGATAGCGGT GGCTGCACGG GCGCAGACAG AGTGGCGGCG

     14650      14660      14670      14680      14690      14700
TCGCCCGATG TCGCGCGGCT TGTTATACGC TAGCCCGTCG CCGCCTCGGG GCACGGTGCC
AGCGGGCTAC AGCGCGCCGA ACAATATGCG ATCGGGCAGC GGCGGAGCCC CGTGCCACGG

     14710      14720      14730      14740      14750      14760
CTCCTACCCA CGTAACTTCC TCCGTGACTT AAAGTCGCGT GTGGTAGATC TCCTGCTCCG
GAGGATGGGT GCATTGAAGG AGGCACTGAA TTTCAGCGCA CACCATCTAG AGGACGAGGC

     14770      14780      14790      14800      14810      14820
TGGACGAACC GTCCGGCAGG ATAGCGGTTA AGGATTCGGT GCTAAGGCCG TGTCGCCAAC
ACCTGCTTGG CAGGCCGTCC TATCGCCAAT TCCTAAGCCA CGATTCCGGC ACAGCGGTTG

     14830      14840      14850      14860      14870      14880
GTCGAATGCT ACGTTGCAAC AGCTTCGACG GACGGCCATC CCCTCTCTCA TCGCAATAAT
CAGCTTACGA TGCAACGTTG TCGAAGCTGC CTGCCGGTAG GGGAGAGAGT AGCGTTATTA

     14890      14900      14910      14920      14930      14940
AAAACACCAG CAGCGCGCAC GACGCGATCA CGGTGACACC CATGATTAGA CCCACGCAGA
TTTTGTGGTC GTCGCGCGTG CTGCGCTAGT GCCACTGTGG GTACTAATCT GGGTGCGTCT

     14950      14960      14970      14980      14990      15000
TAGCCAGCCC CGCTAGCGTA TCTAGCGCCA TCCCGTTCGC TCCCGTTGTC TCCTGAGCGA
ATCGGTCGGG GCGATCGCAT AGATCGCGGT AGGGCAAGCG AGGGCAACAG AGGACTCGCT

     15010      15020      15030      15040      15050      15060
AGCAACTTCT CGGTCCCCGT TTTCAACAGT TTTTGTTTCC TTCTCCGCGA CTAGATGTTA
TCGTTGAAGA GCCAGGGGCA AAAGTTGTCA AAAACAAAGG AAGAGGCGCT GATCTACAAT

     15070      15080      15090      15100      15110      15120
ACGCCCGCGG TCTTTCCGGC CGTGCTCTAC CTCCTGGCGC TTGTCGTCTG GGTTGAGATG
TGCGGGCGCC AGAAAGGCCG GCACGAGATG GAGGACCGCG AACAGCAGAC CCAACTCTAC

     15130      15140      15150      15160      15170      15180
TTCTGCCTCG TCGCCGTAGC CGTCGTCGAG CGCGAGATCG CCTGGGCGCT GCTGCTGCGG
AAGACGGAGC AGCGGCATCG GCAGCAGCTC GCGCTCTAGC GGACCCGCGA CGACGACGCC

     15190      15200      15210      15220      15230      15240
ATGCTGGTCG TTGGCCTGAT GGTGGAAGTC GGCGCCGCCG CCGCTTGGAC CTTCGTGCGT
TACGACCAGC AACCGGACTA CCACCTTCAG CCGCGGCGGC GGCGAACCTG GAAGCACGCA

     15250      15260      15270      15280      15290      15300
TGTCTTGCCT ATCAGCGCTC CTTCCCCGTG CTTACGGCCT TCCCCTGAAA CCCACGTTAA
ACAGAACGGA TAGTCGCGAG GAAGGGGCAC GAATGCCGGA AGGGGACTTT GGGTGCAATT
```

Fig. 10

```
      15310       15320       15330       15340       15350       15360
CCGACCGTCC CAAAAACGCC GGTGTTAACA CAGGAAAAAA AGAAACCACG CAGGAACCGC
GGCTGGCAGG GTTTTTGCGG CCACAATTGT GTCCTTTTTT TCTTTGGTGC GTCCTTGGCG

      15370       15380       15390       15400       15410       15420
GCAGGAACCA CGCGGAACAT GGGACACTAT CTGGAAATCC TGTTCAACGT CATCGTCTTC
CGTCCTTGGT GCGCCTTGTA CCCTGTGATA GACCTTTAGG ACAAGTTGCA GTAGCAGAAG

      15430       15440       15450       15460       15470       15480
ACTCTGCTGC TCGGCGTCAT GGTCAGTATC GTCGCTTGGT ACTTCACGTG AACCACCGTC
TGAGACGACG AGCCGCAGTA CCAGTCATAG CAGCGAACCA TGAAGTGCAC TTGGTGGCAG

      15490       15500       15510       15520       15530       15540
GTCCCGGTTT AAAAACCATC ATCGACGGCC GTTATAAAGC CACCCGGACA CGCGCCGCGG
CAGGGCCAAA TTTTTGGTAG TAGCTGCCGG CAATATTTCG GTGGGCCTGT GCGCGGCGCC

      15550       15560       15570       15580       15590       15600
CACTTGCCTA CGGCGCTGCT TCAGGGAAAC TCCTCTTCCT TCTGCTCTTC CTCCTTCACC
GTGAACGGAT GCCGCGACGA AGTCCCTTTG AGGAGAAGGA AGACGAGAAG GAGGAAGTGG

      15610       15620       15630       15640       15650       15660
GCAGGGATCG TTTCCCTCGA CCAGGGACTC GCCGAAGCAA CCGCCGGAGC AACCTGGAGG
CGTCCCTAGC AAAGGGAGCT GGTCCCTGAG CGGCTTCGTT GGCGGCCTCG TTGGACCTCC

      15670       15680       15690       15700       15710       15720
AGTCGCGGCA TGACGGCGCC CAAGTGTGTC ACCACCAGTA CTTATCTGGT CAAGACCAAG
TCAGCGCCGT ACTGCCGCGG GTTCACACAG TGGTGGTCAT GAATAGACCA GTTCTGGTTC

      15730       15740       15750       15760       15770       15780
GAACAGCCCT GGTGGCCCGA CAACGCCATC AGGAGATGGT GGATCAGTGT TGCTATCGTC
CTTGTCGGGA CCACCGGGCT GTTGCGGTAG TCCTCTACCA CCTAGTCACA ACGATAGCAG

      15790       15800       15810       15820       15830       15840
ATCTTCATCG GAGTCTGTCT GGTGGCCCTG ATGTACTTTA CGCAGCAGCA GGCACGCAGC
TAGAAGTAGC CTCAGACAGA CCACCGGGAC TACATGAAAT GCGTCGTCGT CCGTGCGTCG

      15850       15860       15870       15880       15890       15900
GGGAGCAGCA GCGGCTAGAC AAGTCTCTGG CGGCTACAGC TCCAAGCGCC GTAGCCGGGC
CCCTCGTCGT CGCCGATCTG TTCAGAGACC GCCGATGTCG AGGTTCGCGG CATCGGCCCG

      15910       15920       15930       15940       15950       15960
CGCCTGCCGA TCGCGACGTC GTGGACCATC GAACAGAGAC TCACGCGTAC GAGACCCCGA
GCGGACGGCT AGCGCTGCAG CACCTGGTAG CTTGTCTCTG AGTGCGCATG CTCTGGGGCT

      15970       15980       15990       16000       16010       16020
GGTACGCCAC GCGGTGCCTA ACGCGGTATA CCACACCCGT ACGGTCTGCA GTGCGGCGTA
CCATGCGGTG CGCCACGGAT TGCGCCATAT GGTGTGGGCA TGCCAGACGT CACGCCGCAT

      16030       16040       16050       16060       16070       16080
CAACGTGTGG AAAACGCGTT GCGTCGCAGA GTCCGCCACG TTCCTGTCTT GTCGCTCCCC
GTTGCACACC TTTTGCGCAA CGCAGCGTCT CAGGCGGTGC AAGGACAGAA CAGCGAGGGG

      16090       16100       16110       16120       16130       16140
AATCGTCTCC CGCACACCCC CCGCGACACC CAGAGGGCGG GTGAGCCAAG TATTCTTAAG
TTAGCAGAGG GCGTGTGGGG GGCGCTGTGG GTCTCCCGCC CACTCGGTTC ATAAGAATTC

      16150       16160       16170       16180       16190       16200
GCCGTTCTTT GTTCCATAGC CCATAAATTG TTGATTCCGG AGCTCGTTGG CGCGGAAATA
CGGCAAGAAA CAAGGTATCG GGTATTTAAC AACTAAGGCC TCGAGCAACC GCGCCTTTAT

      16210       16220       16230       16240       16250       16260
GCCGGATAAG GGGAGCAACA ACCGTTGGCG AAAGCCGTCC CGCTCATTCA GTCCGGGTTT
CGGCCTATTC CCCTCGTTGT TGGCAACCGC TTTCGGCAGG GCGAGTAAGT CAGGCCCAAA

      16270       16280       16290       16300       16310       16320
CGCGTCCAGT CGGACGTGTG ACCGTTGGGC AACGGAACGG CGTTTCACTG CCAAAATCGT
GCGCAGGTCA GCCTGCACAC TGGCAACCCG TTGCCTTGCC GCAAAGTGAC GGTTTTAGCA
```

Fig. 1P

```
      16330       16340       16350       16360       16370       16380
ATCGGGTAGT GTACGAGACG TCGGCGGTGC AGAATGCGAC TCGCGGCGTA GCTCGCCGTC
TAGCCCATCA CATGCTCTGC AGCCGCCACG TCTTACGCTG AGCGCCGCAT CGAGCGGCAG

      16390       16400       16410       16420       16430       16440
GCTATGCGGC TCGTCGCCGT GTGGCGCGGC CTGGCCGGCT GTCTGCGTCC AGATCTGTTG
CGATACGCCG AGCAGCGGCA CACCGCGCCG GACCGGCCGA CAGACGCAGG TCTAGACAAC

      16450       16460       16470       16480       16490       16500
GCCTTTTGGT TCCTCTGGCT GCTGCTGCGT GTGTGCTTTG GTAGACGCGG TGGCAGTTTG
CGGAAAACCA AGGAGACCGA CGACGACGCA CACACGAAAC CATCTGCGCC ACCGTCAAAC

      16510       16520       16530       16540       16550       16560
CGGTCTGCGG TAAGTGAGGA TGTCGCCGAG CAAACGCACT TGCGGCGCGT GGGCGGCACG
GCCAGACGCC ATTCACTCCT ACAGCGGCTC GTTTGCGTGA ACGCCGCGCA CCCGCCGTGC

      16570       16580       16590       16600       16610       16620
CGTGTCATTG TAGGTTCGTT GCCAGATGGC AAGTGCTGTC AACAGCAGGC GTTGTGGGCG
GCACAGTAAC ATCCAAGCAA CGGTCTACCG TTCACGACAG TTGTCGTCCG CAACACCCGC

      16630       16640       16650       16660       16670       16680
GTCGGTGTAT TTTTGTGGGT TGCGGTGAGA GTCGGCACTC GGTGTTTTGT GAGTCATCTC
CAGCCACATA AAAACACCCA ACGCCACTCT CAGCCGTGAG CCACAAAACA CTCAGTAGAG

      16690       16700       16710       16720       16730       16740
AACTATCTGT GTTGCTTTGA GCAGCGTCCA GAACAGCGAC GCGACTTTGG GGATGGCCTC
TTGATAGACA CAACGAAACT CGTCGCAGGT CTTGTCGCTG CGCTGAAACC CCTACCGGAG

      16750       16760       16770       16780       16790       16800
GTGCTCACCT CCGCGGAGAG CGCCGCCGGA CCTGCTCGTC AGCAGCGAGC TACGCAGACG
CACGAGTGGA GGCGCCTCTC GCGGCGGCCT GGACGAGCAG TCGTCGCTCG ATGCGTCTGC

      16810       16820       16830       16840       16850       16860
GAATATCTGG AGGAGAGTTA CGTGTGTCAC AGGAGAGCGC GGGTCTCCGG CGGTAACGAC
CTTATAGACC TCCTCTCAAT GCACACAGTG TCCTCTCGCG CCCAGAGGCC GCCATTGCTG

      16870       16880       16890       16900       16910       16920
GGCGGTGTCG TCGACACGTG TGCGGCCTGT TGTGCTCTGC GGAAAAGTGC CGGTCTCGGA
CCGCCACAGC AGCTGTGCAC ACGCCGGACA ACACGAGACG CCTTTTCACG GCCAGAGCCT

      16930       16940       16950       16960       16970       16980
GACCGTGGAC GAAAAAGAGA ACGCAGCAGC TACCGCTGGC GGCGGCGGCG TTAATGCAGC
CTGGCACCTG CTTTTTCTCT TGCGTCGTCG ATGGCGACCG CCGCCGCCGC AATTACGTCG

      16990       17000       17010       17020       17030       17040
CGTTGATGTT CGACGTTGTG AGCACTCGGA AACAGCGGTG AGGCAGAAGG TCGATTCTCC
GCAACTACAA GCTGCAACAC TCGTGAGCCT TTGTCGCCAC TCCGTCTTCC AGCTAAGAGG

      17050       17060       17070       17080       17090       17100
AGGGAACGAC AGTCGATGCG TGGTAGCCGC AGCAGGTGAG GTTGGGGCGG ACAACGTGTT
TCCCTTGCTG TCAGCTACGC ACCATCGGCG TCGTCCACTC CAACCCCGCC TGTTGCACAA

      17110       17120       17130       17140       17150       17160
GCGGATTGTG GCGAGAACGT CGTCCTCCCC TTCTTCACCG CCCCACCCAC CCTCGGTTGG
CGCCTAACAC CGCTCTTGCA GCAGGAGGGG AAGAAGTGGC GGGGTGGGTG GGAGCCAACC

      17170       17180       17190       17200       17210       17220
TGTTTCTTTT TTCTTGTGTC CTGCAGATAG TTCCACGGAC AGCGACGGCA AGTCCATAAT
ACAAAGAAAA AAGAACACAG GACGTCTATC AAGGTGCCTG TCGCTGCCGT TCAGGTATTA

      17230       17240       17250       17260       17270       17280
CAGCGGTGTG CAAGTGGTGG AACACGACGA AGATATCATC GCGCCGCAGA GTTTGTGGTG
GTCGCCACAC GTTCACCACC TTGTGCTGCT TCTATAGTAG CGCGGCGTCT CAAACACCAC

      17290       17300       17310       17320       17330       17340
CACGGCGTTC AAGGAAGCCC TCTGGGATGT GGCTCTGTTG GAAGTGCCGC GTTGGGCGTG
GTGCCGCAAG TTCCTTCGGG AGACCCTACA CCGAGACAAC CTTCACGGCG CAACCCGCAC
```

Fig. 1Q

```
     17350      17360      17370      17380      17390      17400
GCAGGGCTGG AAGAGGTGGC GCAACAGCGA GGCCGGGCGT CGATGGAGTG CTGGGTCTGC
CGTCCCGACC TTCTCCACCG CGTTGTCGCT CCGGCCCGCA GCTACCTCAC GACCCAGACG

     17410      17420      17430      17440      17450      17460
GTCGGCTTCC AGCTTGTCTG ACTTGGCGGG CGAGGCCGTT GGAGAATTGG TGGGATCGGT
CAGCCGAAGG TCGAACAGAC TGAACCGCCC GCTCCGGCAA CCTCTTAACC ACCCTAGCCA

     17470      17480      17490      17500      17510      17520
CGTCGCGTAC GTGATCCTTG AACGTCTGTG GTTGGCAGCC AGAGGTTGGG TGTGCGAAAC
GCAGCGCATG CACTAGGAAC TTGCAGACAC CAACCGTCGG TCTCCAACCC ACACGCTTTG

     17530      17540      17550      17560      17570      17580
AGGTGTGGAA GCCGAGGAGG CCATGTCGCG GCGGCGACAG CGCATGCTGT GGCGTATTGT
TCCACACCTT CGGCTCCTCC GGTACAGCGC CGCCGCTGTC GCGTACGACA CCGCATAACA

     17590      17600      17610      17620      17630      17640
TCTCTCGTGG AGGCGACGGC GAATGCAGCA GACGGTGTTC GATGGAGATG GCGTGCGGGG
AGAGAGCACC TCCGCTGCCG CTTACGTCGT CTGCCACAAG CTACCTCTAC CGCACGCCCC

     17650      17660      17670      17680      17690      17700
AAGAAAGCGC CGTGTTGTGA GCAGACGACG TAGGATGCGG GACGTCGGAG CACATGGGCC
TTCTTTCGCG GCACAACACT CGTCTGCTGC ATCCTACGCC CTGCAGCCTC GTGTACCCGG

     17710      17720      17730      17740      17750      17760
ATGTGTGGTG GCAGATGGCG GTGTCCGCTG GTGTCTGCTG CGGCAGTGCA TAGACGAAGC
TACACACCAC CGTCTACCGC CACAGGCGAC CACAGACGAC GCCGTCACGT ATCTGCTTCG

     17770      17780      17790      17800      17810      17820
AACATGTCGC TGTGAAGAGA TAGAGTGTGA GCATAGCTGC ATGCAGCGTT GCGTGTATAA
TTGTACAGCG ACACTTCTCT ATCTCACACT CGTATCGACG TACGTCGCAA CGCACATATT

     17830      17840      17850      17860      17870      17880
GCGGGGGGGA TTAAGACGTT AATAAAGAAT AGCGGCGGTT CTGATAGGGC GACCGCTGAA
CGCCCCCCCT AATTCTGCAA TTATTTCTTA TCGCCGCCAA GACTATCCCG CTGGCGACTT

     17890      17900      17910      17920      17930      17940
GTGAGCTGCG TGTGCGTGTG GTTTGTGGAG TCCCCGCCGC CCCCGGTCCC GTGTCCGCCG
CACTCGACGC ACACGCACAC CAAACACCTC AGGGGCGGCG GGGGCCAGGG CACAGGCGGC

     17950      17960      17970      17980      17990      18000
GCAAAGCCCC CCGGNTCCGC ACACTCCTGG CCGCGCAACC CTCGTCGCTG CAAAAGCCCC
CGTTTCGGGG GGCCNAGGCG TGTGAGGACC GGCGCGTTGG GAGCAGCGAC GTTTTCGGGG

     18010      18020      18030      18040      18050      18060
CCGTCCCCGC ACACCCCCGC GACCGCCGGT CCCGCGAGTC CCCGTCCCCG CCGCAAAAGG
GGCAGGGGCG TGTGGGGGCG CTGGCGGCCA GGGCGCTCAG GGGCAGGGGC GGCGTTTTCC

     18070      18080      18090      18100      18110      18120
CCCCCGTCCT CGCCGCAAAC ACCCCCGTCA CCCCCGTCCC TCAGNCCGGG TCCGCGAGTC
GGGGGCAGGA GCGGCGTTTG TGGGGGCAGT GGGGGCAGGG AGTCNGGCCC AGGCGCTCAG

     18130      18140      18150      18160      18170      18180
CCCGTTCCCA GCGTAATCCC CGTACCCGCA ACGNCCCGGN CCCACCGTCG TCCCGCACAC
GGGCAAGGGT CGCATTAGGG GCATGGGCGT TGCNGGGCCN GGGTGGCAGC AGGGCGTGTG

     18190      18200      18210      18220      18230      18240
CCCCCGTCCC CCAGCCCGGT GCCCAGCGTG CGAAAAAAGC TCCGTCCCTC ACACCCGCAG
GGGGGCAGGG GGTCGGGCCA CGGGTCGCAC GCTTTTTTCG AGGCAGGGAG TGTGGGCGTC

     18250      18260      18270      18280      18290      18300
AAAGATCCCT CAGCGCGGTG AAACCCCGTC CCCAGCGCCG TGCCGCTGAC AAAGACCATG
TTTCTAGGGA GTCGCGCCAC TTTGGGGCAG GGGTCGCGGC ACGGCGACTG TTTCTGGTAC

     18310      18320      18330      18340      18350      18360
GGACGACACG CACAGGCA.. .......... .......... .......... ..........
CCTGCTGTGC GTGTCCGT.. .......... .......... .......... ..........
```

Fig. 1R

Attorney Docket No. AVIR-0032-1

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I declare that:
My residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: ATTENUATION OF CYTOMEGALOVIRUS VIRULENCE the specification of which is attached hereto. I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign applications(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Country	Application No.	Date of Filing	Priority Claimed Under 35 USC 119
			Yes ___ No ___
			Yes ___ No ___

I hereby claim the benefit under Title 35, United States Code 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status
		___ Patented ___ Pending ___ Abandoned
		___ Patented ___ Pending ___ Abandoned

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Tracy J. Dunn Reg. No. 34,587

Send Correspondence to:	Direct Telephone Calls to:
Tracy J. Dunn **AVIRON** **297 Bernardo Avenue** **Mt. View, CA 94043**	Tracy J. Dunn Reg. No. 34,587 650.919.6637 (voice) 650.919.6610 (fax)

Full Name of Inventor 1	Last Name KEMBLE	First Name George	Middle Name or Initial W.	
Residence & Citizenship	City Fremont	State/Foreign Country CALIFORNIA	Country of Citizenship USA	
Post Office Address	Post Office Address 44190 Boitano Avenue	City Fremont	State/Country CALIFORNIA	Zip Code 94539
Full Name of Inventor 2	Last Name DUKE	First Name Gregory	Middle Name or Initial M.	
Residence & Citizenship	City Palo Alto	State/Foreign Country CALIFORNIA	Country of Citizenship USA	
Post Office Address	Post Office Address 659 Forest Avenue	City Palo Alto	State/Country CALIFORNIA	Zip Code 94301
Full Name of Inventor 3	Last Name SPAETE	First Name Richard	Middle Name or Initial R.	
Residence & Citizenship	City Redwood City	State/Foreign Country California	Country of Citizenship USA	
Post Office Address	Post Office Address 720 Lakemead Way.	City Redwood City	State/Country CALIFORNIA	Zip Code 94062

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1	Signature of Inventor 2	Signature of Inventor 3
GEORGE W. KEMBLE	GREGORY M. DUKE	RICHARD R.SPAETE
Date	Date	Date